TOPICS IN THE STATISTICAL ASPECTS OF SIMULATION

A Thesis
Presented to
The Academic Faculty

by

Joshua L. McDonald

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial & Systems Engineering

Georgia Institute of Technology
December 2015

TOPICS IN THE STATISTICAL ASPECTS OF SIMULATION

Approved by:

Professor David Goldsman, Advisor
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

Professor Christos Alexopoulos
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

Professor Laurie Garrow
School of Civil and Environmental Engineering
Georgia Institute of Technology

Professor Paul Griffin
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

Professor Brani Vidakovic
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

Date Approved: 11 August 2015

For those who didn't make it, and those still working towards it

ACKNOWLEDGEMENTS

Thanks to the faculty who provided guidance, the staff who smoothed the path, my committee for their willingness to serve, and my friends and family who have been so supportive throughout this endeavor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In the first part of the thesis, we apply various variance reduction techniques to the estimation of Asian averages and options and propose an easy-to-use quasi-Monte Carlo method that can provide significant variance reductions with minimal increases in computational time. We have also extended these techniques to estimate higher moments of the Asians averages. In the second part, we then use these estimated moments to efficiently implement Gram–Charlier based estimators for probability density functions of Asian averages and options. Finally, in the third part, we investigate a ranking and selection application that uses post hoc analysis to determine how the circumstances of procedure termination affect the probability of correct selection.

CHAPTER I

INTRODUCTION

1.1 Overview

This thesis concerns various problems arising in the analysis of certain stochastic systems. Simulation is our primary analysis tool. In the first portion of the thesis, we are interested in the efficient simulation of stochastic processes that arise in financial engineering applications. In particular, we use simulation to estimate performance characteristics of Asian averages, for example, the expected value or quantiles of a stock's average over a given time period; we also consider the associated options on those averages.

In a separate line of research, the thesis also studies new performance characteristics related to statistical ranking and selection procedures. In such procedures, we are interested in determining which of a number of populations or alternatives is the "best", where that term depends on the context of the problem. Ranking and selection procedures are typically regarded as experimental designs, where the overall probability of making a correct selection is, informally, determined prior to the start of experimentation. The thesis looks at the problem of estimating the probability of correct selection after experimentation has taken place.

Additional general remarks on our research, as well as organizational details, follow in the subsections below.

1.1.1 Efficient Simulation of Asian Averages

We first study methods that enable the efficient simulation of various random processes that arise in financial engineering. It is well known that one can model an underlying stock's price via geometric Brownian motion (GBM), which is very easy to simulate by exponentiating the sum of appropriate independent and identically distributed (i.i.d.) normal increments. In this part of the thesis, we focus on estimating quantities related to Asian averages of GBM processes, where an Asian is simply the average of the stock prices taken over a certain time period.

Asian averages come in many forms — additive, geometric, and harmonic; we will concentrate mainly on additive averages. Specifically, we are interested in the probability density function (p.d.f.) of these averages, but we will settle for their first few moments and quantiles. We also study the analogous quantities for options based on the Asian averages.

The main trick we use is to employ quasi-Monte Carlo (QMC) simulation to manipulate the underlying Brownian motion increments in such a way as to effect variance reduction. In order to implement QMC, we generate multiple, correlated sample paths of the GBM for a single set of i.i.d. normal random variables. This method saves on the expense of generating a prohibitive number of normals, while incurring only a modest cost due to additional computations. If the sample paths are correlated in a certain way, it is possible to reduce the variance of moment estimators, compared to estimates that do not incorporate the correlated sample paths. So instead of conducting multiple independent simulation replications of the Asian averages, we can use techniques such as antithetic variates to reduce the overall computational burden, that is, by reducing the number of necessary replications to obtain unbiased estimators with comparable variances. This portion of the thesis shows how to structure and evaluate our variance reduction methods for use on Asian averages and then on the associated Asian options.

1.1.2 Gram–Charlier Pricing of Asian Averages and Options

The second topic of the thesis builds on the previous subject matter. With the various sample paths in hand from §1.1.1, we can approximate the p.d.f. of the Asian average either by (i) using the sample paths to produce a histogram of Asian average values, or (ii) using the moment estimators in conjunction with a Gram–Charlier (GC)-type p.d.f. approximation. This latter GC work is the topic of the second part of the thesis.

In either case (i) or (ii), once we have the approximate Asian average p.d.f. at our disposal, we can then price various derivatives based on the Asian average, for instance, call and put options. We do so with sample estimators that have lower variance than the analogous naive Monte Carlo estimators. Similarly, we can estimate the so-called "Greeks" associated with the options, that is, the sensitivities of the options with respect to such parameters as stock prices, volatility, etc. We study examples in which the proposed methodology is expected to perform well and examples where the approximations break down — not surprisingly, when the volatility σ driving the underlying Brownian motion starts to get too high.

We also discuss several generalizations of the methodology, including their use in situations where the exponential Brownian motion of GBM is replaced by a Lévy process, that is, a stationary stochastic process having independent increments that are non-normal.

1.1.3 Conditional Probability of Correct Selection for Ranking and Selection Procedures

The third thesis topic belongs to the general area of ranking and selection (R&S) theory. R&S seeks to select the best of a number of competing populations, based on various sampling schemes, subject to a constraint that guarantees an achieved probability of correct selection ($P\{\text{CS}\}$). In the usual formulations of R&S problems, a guaranteed lower bound on $P\{\text{CS}\}$ is specified a priori. We, instead, will study the conditional $P\{\text{CS}\}$ after sampling has concluded — which may be substantially different than the a priori version. By taking into account both when and how the R&S procedure terminates, the experimenter can compute or estimate the conditional $P\{\text{CS}\}$ given the observed scenario. This can be especially useful for situations in which there is a significant penalty for making an incorrect decision.

This approach certainly makes sense from a practical point of view. For example, suppose that two competing drugs are tested on two samples of people, and further suppose that a certain R&S procedure dictates that in order to achieve an a priori $P\{\text{CS}\}$ of 0.9, (i) each sample must be tested on 100 people, and (ii) whichever drug achieves the most successes on its sample of 100 people will be declared the winner, i.e., the most efficacious. It stands to reason that we would be much more confident about our posterior $P\{\text{CS}\}$ if drug A were to win over drug B by a margin of 80 to 20, than by a margin of 51 to 49 — the latter of which amounts to little more than a coin toss.

We will examine this type of scenario with various popular, well-known R&S procedures in order to study the effects of procedure termination on the conditional probability of correct selection.

1.2 Organization of the Thesis

Chapter 2 is concerned with the efficient simulation of Asian averages. Chapter 3 builds on the previous work by using the Asian average's estimated moments to approximate the p.d.f. of the average by a Gram–Charlier distribution, which is subsequently used to price the associated Asian options. Chapter 4 discusses the conditional probability of correct selection that is achieved after certain ranking and selection procedures terminate. Finally, Chapter 5 gives conclusions and outlines additional work that will be interesting going forward.

CHAPTER II

EFFICIENT SIMULATION OF ASIAN AVERAGES

This chapter studies the efficient pricing of certain averages of underlying equity prices. In particular, we develop several variance reduction techniques to estimate moments of Asian averages and then Asian options, which are important quantities in financial engineering.

The chapter proceeds as follows. §2.1 gives introductory material on the underlying stock price model, along with basic definitions involving Asian averages and their associated options. §2.2 deals with geometric Asian averages and options, the latter of which can be handled in a straightforward manner via classical methods. §2.3 concerns the analogous problems involving arithmetic averages; these are more difficult to analyze than their geometric brethren. §2.4 discusses a variety of extensions.

2.1 Background

We will consider a stock whose value over time $(S(t),\, t \geq 0)$ is driven by a standard Brownian motion (BM) process $(\mathcal{W}(t),\, t \geq 0)$ [8]. It is well-known that Brownian motion possesses a number of useful properties, among them:

1. $\mathcal{W}(0) = 0$.

2. The process $(\mathcal{W}(t),\, t \geq 0)$ is Gaussian. In particular, all joint distributions are multivariate normal; and specifically, $\mathcal{W}(t) \sim \mathrm{Nor}(0, t)$ for all $t \geq 0$.

3. The process $(\mathcal{W}(t),\, t \geq 0)$ is stationary. In particular, $\mathcal{W}(t+h) - \mathcal{W}(t) \sim \mathrm{Nor}(0, h)$ for all $t \geq 0$ and $h \geq 0$.

4. The process $(\mathcal{W}(t),\, t \geq 0)$ has independent increments. In particular, for $0 \leq$

$a \leq b \leq c \leq d$, we have that $\mathcal{W}(d) - \mathcal{W}(c)$ is independent of $\mathcal{W}(b) - \mathcal{W}(a)$.

5. For $0 \leq a \leq b$, we have $\mathrm{Cov}(\mathcal{W}(a), \mathcal{W}(b)) = a$.

The stock price itself evolves over time t via the classic geometric Brownian motion (GBM) process,

$$S(t) \;\equiv\; s_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma\mathcal{W}(t)\right\} \;\sim\; s_0\;\exp\left\{\mathrm{Nor}\left(\left(\mu - \frac{\sigma^2}{2}\right)t,\, \sigma^2 t\right)\right\}, \quad t \geq 0, \tag{1}$$

where $s_0 \equiv S(0)$ is the (known) initial price; and $\mu \in \mathbb{R}$ and $\sigma > 0$ respectively represent the natural drift of the stock (hopefully positive) and the natural random volatility due to the underlying Brownian motion process. For the remainder of this chapter we set $\mu = r$, where r is the risk-free interest rate, for example, the interest rate offered by U.S. Treasury bonds.

We will study properties associated with the average stock price over the time interval $[0, T]$, where T is specified in advance, and where the term "average" can be defined in several ways. First of all, one can take either an arithmetic or a geometric (or even a harmonic) average; and second, these averages can be based on either continuous or discrete sampling [8]. In any case, these averages are generally referred to as *Asian averages*.

We denote the discretely and continuously sampled arithmetic averages by

$$A_m \;\equiv\; \frac{1}{m}\sum_{i=1}^{m} S(t_i) \quad \text{and} \quad A \;\equiv\; \frac{1}{T}\int_0^T S(t)\,\mathrm{d}t,$$

respectively, where the number of sampling points m is specified beforehand and for ease of exposition, we henceforth set $t_i \equiv iT/m$, $i = 1, 2, \ldots, m$. By the continuous mapping theorem [5], it is easy to show that $A_m \Rightarrow A$ as $m \to \infty$, where $\Rightarrow$ denotes weak convergence. Similarly, the discretely and continuously sampled geometric averages are

$$G_m \;\equiv\; \left(\prod_{i=1}^{m} S(t_i)\right)^{1/m} \quad \text{and} \quad G \;\equiv\; \exp\left(\frac{1}{T}\int_0^T \ell\mathrm{n}(S(t))\,\mathrm{d}t\right),$$

respectively; and $G_m \Rightarrow G$ as $m \to \infty$.

One of this chapter's goals is to obtain properties of the distributions of A_m, A, G_m, and G, or at least their low-order moments. In addition, we can consider an Asian call option at expiry time T for strike price k, i.e., $C_X \equiv (X - k)^+$, where X is one of A_m, A, G_m, and G, and $(y)^+ \equiv \max\{y, 0\}$. Such a call gives us the right, but not the obligation, to buy a share of the underlying stock at price k at time T. The call's fair value at expiry, discounted back to time 0, is

$$c_X \equiv e^{-rT}\mathrm{E}[C_X].$$

It is well known that the problem of obtaining results such as those outlined above is relatively easier for geometric averages than for arithmetic averages. We shall discuss these cases in §§2.2 and 2.3, respectively, the latter of which gives the main content of this chapter.

2.2 Geometric Average of the Equity Price

§2.2.1 reviews basic results on the geometric average of an equity price, and §2.2.2 does the same for the corresponding call option.

2.2.1 Basic Results

We begin with some simple results on the geometric average of the equity price over the time interval $[0, T]$. For the discretely monitored version, we use Equation (1) with $t_i = iT/m$, $i = 1, 2, \ldots, m$, and the facts that $\mathrm{Cov}[\mathcal{W}(u), \mathcal{W}(v)] = \min(u, v)$ and

$$\mathrm{Var}\left[\sum_{i=1}^{m} \mathcal{W}(t_i)\right] = \frac{T}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}\min(i, j) = \frac{T(2m+1)(m+1)}{6}$$

to obtain

$$\begin{aligned} G_m &= s_0 \exp\left\{\frac{1}{m}\sum_{i=1}^{m}\left(\left(r - \frac{\sigma^2}{2}\right)t_i + \sigma\mathcal{W}(t_i)\right)\right\} \\ &\sim s_0 \exp\left\{\mathrm{Nor}\left(\left(r - \frac{\sigma^2}{2}\right)T\alpha_m,\, \sigma^2 T\beta_m\right)\right\}, \qquad (2) \end{aligned}$$

where $\alpha_m \equiv (m+1)/(2m)$ and $\beta_m \equiv (2m+1)(m+1)/(6m^2)$ for $m \geq 1$.

Similarly, for the continuously monitored version, we can either take the limit as $m \to \infty$ in (2) or instead use (1) and the fact that

$$\mathrm{Var}\left[\int_0^T \mathcal{W}(t)\,\mathrm{d}t\right] = \int_0^T \int_0^T \min(u,v)\,\mathrm{d}u\,\mathrm{d}v = \frac{T^3}{3}$$

to obtain

$$\begin{aligned} G &= s_0 \exp\left\{\frac{1}{T}\int_0^T \left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma\mathcal{W}(t)\right)\mathrm{d}t\right\} \\ &\sim s_0 \exp\left\{\mathrm{Nor}\left(\left(r - \frac{\sigma^2}{2}\right)\frac{T}{2}, \frac{\sigma^2 T}{3}\right)\right\}. \end{aligned} \tag{3}$$

Then by elementary properties of the lognormal distribution,

$$\begin{aligned} \mathrm{E}[G_m] &= s_0 \exp\left\{\left(r - \frac{\sigma^2}{2}\right)T\alpha_m + \frac{\sigma^2 T\beta_m}{2}\right\} \\ &\to s_0 \exp\left\{\frac{T}{12}\left(6r - \sigma^2\right)\right\} = \mathrm{E}[G] \quad \text{as } m \to \infty. \end{aligned} \tag{4}$$

Similarly,

$$\begin{aligned} \mathrm{Var}[G_m] &= s_0^2\left[\exp\{\sigma^2 T\beta_m\} - 1\right]\exp\left\{2T\alpha_m\left(r - \frac{\sigma^2}{6} + \frac{\sigma^2}{6m}\right)\right\} \\ &\to s_0^2\left[\exp\left\{\frac{\sigma^2 T}{3}\right\} - 1\right]\exp\left\{\frac{T}{6}\left(6r - \sigma^2\right)\right\} = \mathrm{Var}[G] \quad \text{as } m \to \infty. \end{aligned}$$

2.2.2 Contingent Claims Based on the Geometric Average

The Asian call option value $c_G = e^{-rT}\mathrm{E}[C_G]$ based on the continuous geometric average of GBM has a closed form [10]. The result also extends to the discrete case $c_{G_m} = e^{-rT}\mathrm{E}[C_{G_m}]$ as follows.

Proposition 1 Let $\Phi(\cdot)$ denote the standard normal cumulative distribution function (c.d.f.). Then

$$c_{G_m} = s_0 \exp\left\{-r(1-\alpha_m)T - (\alpha_m - \beta_m)\frac{\sigma^2 T}{2}\right\}\Phi(z_{G_m}^+) - k\,\mathrm{e}^{-rT}\,\Phi(z_{G_m}^-) \tag{5}$$

$$\to s_0\,\mathrm{e}^{-(r+\frac{\sigma^2}{6})\frac{T}{2}}\,\Phi(z_G^+) - k\,\mathrm{e}^{-rT}\,\Phi(z_G^-) = c_G, \quad \text{as } m \to \infty, \tag{6}$$

where

$$z_{G_m}^+ \;\equiv\; \frac{\ell\text{n}\left(\frac{s_0}{k}\right) + r\alpha_m T + (\beta_m - \frac{\alpha_m}{2})\sigma^2 T}{\sigma\sqrt{\beta_m T}} \;\to\; \frac{\ell\text{n}\left(\frac{s_0}{k}\right) + (r + \frac{\sigma^2}{6})\frac{T}{2}}{\sigma\sqrt{\frac{T}{3}}} \;\equiv\; z_G^+, \quad \text{and}$$

$$z_{G_m}^- \;\equiv\; \frac{\ell\text{n}\left(\frac{s_0}{k}\right) + (r - \frac{\sigma^2}{2})\alpha_m T}{\sigma\sqrt{\beta_m T}} \;\to\; \frac{\ell\text{n}\left(\frac{s_0}{k}\right) + (r - \frac{\sigma^2}{2})\frac{T}{2}}{\sigma\sqrt{\frac{T}{3}}} \;\equiv\; z_-^G \quad \text{as } m \to \infty. \quad (7)$$

Equations (5) and (6) are versions of the classic Black–Scholes–Merton (BSM) formula [6]. The standard European vanilla call option is a special case obtained by setting $m = 1$ (i.e., no averaging).

2.3 Arithmetic Average of the Equity Price

In this section, we study properties of the arithmetic average of the equity price. These properties will be useful when we eventually price the associated option on the arithmetic average.

To get things going, define the constants

$$\phi \;\equiv\; \left(r - \frac{\sigma^2}{2}\right)\frac{T}{m}, \quad \tau^2 \;\equiv\; \frac{\sigma^2 T}{m}, \quad \text{and} \quad \theta \;\equiv\; \phi + \frac{\tau^2}{2} \;=\; \frac{rT}{m},$$

as well as the increments

$$X_j \;\equiv\; \sigma\left[\mathcal{W}\left(\frac{jT}{m}\right) - \mathcal{W}\left(\frac{(j-1)T}{m}\right)\right], \quad j = 1, 2, \ldots, m.$$

Then we can express the arithmetic average as

$$\begin{aligned} A_m^+ \;&\equiv\; \frac{1}{m}\sum_{i=1}^{m} S\left(\frac{iT}{m}\right) \\ &=\; \frac{s_0}{m}\sum_{i=1}^{m}\exp\left\{\phi i + \sigma\mathcal{W}\left(\frac{iT}{m}\right)\right\} \\ &=\; \frac{s_0}{m}\sum_{i=1}^{m}\exp\left\{\phi i + \sum_{j=1}^{i} X_j\right\}, \end{aligned} \quad (8)$$

where, by independent increments of Brownian motion, $X_1, X_2, \ldots, X_m$ are independent and identically distributed (i.i.d.) $\text{Nor}(0, \tau^2)$. Using elementary properties of the lognormal distribution, we can immediately calculate the expected value of A_m^+,

$$\mathrm{E}[A_m^+] \;=\; \frac{s_0}{m}\sum_{i=1}^{m}\mathrm{E}\big[e^{\mathrm{Nor}(\phi i,\tau^2 i)}\big] \;=\; \frac{s_0}{m}\sum_{i=1}^{m} e^{\theta i} \;=\; \frac{s_0 e^{\theta}(e^{m\theta}-1)}{m(e^{\theta}-1)} \;\equiv\; f_m(\theta). \qquad (9)$$

Even though we can calculate $\mathrm{E}[A_m^+]$ exactly, it is still of interest to study Monte Carlo (MC) methods for estimating $\mathrm{E}[A_m^+]$. The reason is that MC methods that turn out to be efficient with respect to the estimation of $\mathrm{E}[A_m^+]$ will likely be efficient for less-tractable cases, for example, (i) higher-order moments $\mathrm{E}[(A_m^+)^k]$ with $k > 1$, and (ii) driving processes whose increments $X_1, X_2, \ldots, X_m$ are not necessarily normal.

Thus, for now, our goal is to come up with unbiased, low-variance estimators for $\mathrm{E}[A_m^+]$. §2.3.1 discusses the "naive" MC estimator for $\mathrm{E}[A_m^+]$, which is simply based on a series of i.i.d. simulation replications of A_m^+. §2.3.2 defines an antithetic estimator, which improves performance by combining the naive estimator with a negatively correlated version of that estimator. §2.3.3 generalizes the discussion of antithetics by obtaining variance and covariance expressions for a richer class of estimators. §2.3.4 defines what we refer to as the full quasi-Monte Carlo estimator for $\mathrm{E}[A_m^+]$ and deals with implementation issues. §2.3.5 is concerned with the estimation of higher-order moments, $\mathrm{E}[(A_m^+)^k]$. Finally, §2.3.6 outlines our game plan for evaluating the value of an Asian option based on the arithmetic average.

2.3.1 Naive Monte Carlo Estimation of $\mathbf{E}[A_m^+]$

We can of course estimate $\mathrm{E}[A_m^+]$ by simulating realizations of the normal increments $X_1, X_2, \ldots, X_m$ to generate i.i.d. realizations of A_m^+. If we denote n such i.i.d. realizations of A_m^+ by $A_{m,1}^+, A_{m,2}^+, \ldots, A_{m,n}^+$, then the *naive* MC estimator of $\mathrm{E}[A_m^+]$ is simply the sample mean $\bar{A}^+ \equiv \frac{1}{n}\sum_{i=1}^{n} A_{m,i}^+$. By construction, $\bar{A}^+$ is unbiased for $\mathrm{E}[A_m^+]$ and has variance $\mathrm{Var}[\bar{A}^+] = \mathrm{Var}[A_{m,1}^+]/n$. The standard error of $\bar{A}^+$ as an estimator for $\mathrm{E}[A_m^+]$ is the estimated standard deviation of $\bar{A}^+$, i.e.,

$$\mathrm{s.e.}[\bar{A}^+] \;\equiv\; \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(A_{m,i}^+ - \bar{A}^+)^2}, \qquad (10)$$

the sample standard deviation of $A^+_{m,1}, A^+_{m,2}, \ldots, A^+_{m,n}$ divided by $\sqrt{n}$. The standard errors for other estimators will be defined similarly in the sequel.

2.3.2 Antithetic Estimation of $\mathrm{E}[A^+_m]$

With only a little additional work, we can also generate realizations of what we shall call the identically distributed *antithetic* version of the arithmetic average,

$$A^-_m \equiv \frac{s_0}{m}\sum_{i=1}^{m}\exp\left\{\phi i - \sum_{j=1}^{i} X_j\right\}, \tag{11}$$

which merely changes the signs of the original i.i.d. $\mathrm{Nor}(0,\tau^2)$ increments $X_1, X_2, \ldots, X_m$. By symmetry of the normal distribution, A^-_m, like A^+_m, is unbiased for $\mathrm{E}[A^+_m]$. Moreover, one would intuitively think that A^+_m and A^-_m are negatively correlated.

Similar to the naive estimator, we can easily generate i.i.d. realizations of A^-_m. If we denote n such i.i.d. antithetic realizations of A^-_m by $A^-_{m,1}, A^-_{m,2}, \ldots, A^-_{m,n}$, then another MC estimator of $\mathrm{E}[A^+_m]$ is given by the complementary sample mean $\bar{A}^- \equiv \frac{1}{n}\sum_{i=1}^{n} A^-_{m,i}$. By construction and symmetry of the normal increments, $\bar{A}^-$ is unbiased for $\mathrm{E}[A^+_m]$ and has variance $\mathrm{Var}[\bar{A}^-] = \mathrm{Var}[\bar{A}^+] = \mathrm{Var}[A^+_{m,1}]/n$.

Let us now put our naive A^+_m and antithetic A^-_m handicraft together. If we define the average of the two versions of the arithmetic average of the equity by $\bar{A}_m \equiv (A^+_m + A^-_m)/2$ and note that $\mathrm{E}[\bar{A}_m] = \mathrm{E}[A^+_m]$, we see that $\bar{A}_m$ is unbiased for $\mathrm{E}[A^+_m]$. In addition,

$$\mathrm{Var}(\bar{A}_m) = \frac{\mathrm{Var}(A^+_m) + \mathrm{Var}(A^-_m) + 2\mathrm{Cov}(A^+_m, A^-_m)}{4} = \frac{\mathrm{Var}(A^+_m) + \mathrm{Cov}(A^+_m, A^-_m)}{2}.$$

If, as we hope, $\mathrm{Cov}(A^+_m, A^-_m)$ is small (or, better yet, negative), then we will have $\mathrm{Var}(\bar{A}_m) \leq \mathrm{Var}(A^+_m)$.

In order to compare this antithetic methodology with that of naive MC, suppose that we again have on hand n i.i.d. realizations of A^+_m denoted by $A^+_{m,1}, A^+_{m,2}, \ldots, A^+_{m,n}$, as well as the n corresponding i.i.d. antithetic realizations of A^-_m denoted by

$A^-_{m,1}, A^-_{m,2}, \ldots, A^-_{m,n}$. The *pairs* $(A^+_{m,1}, A^-_{m,1}), (A^+_{m,2}, A^-_{m,2}), \ldots, (A^+_{m,n}, A^-_{m,n})$ are themselves bivariate i.i.d.; but for any given pair i, the naive iteration and its antithetic counterpart, $A^+_{m,i}$ and $A^-_{m,i}$, are not independent. As before, we define $\bar{A}^+ \equiv \frac{1}{n}\sum_{i=1}^n A^+_{m,i}$ and $\bar{A}^- \equiv \frac{1}{n}\sum_{i=1}^n A^-_{m,i}$. Then what is commonly referred to as the *antithetic* estimator of $\mathrm{E}[A^+_m]$ based on the n pairs $(A^+_{m,i}, A^-_{m,i})$, $i = 1, 2, \ldots, n$, is $\bar{A} \equiv (\bar{A}^+ + \bar{A}^-)/2$.

Similar to the above discussion, it is easy to see that $\bar{A}$ is unbiased for $\mathrm{E}[A^+_m]$, and

$$\mathrm{Var}(\bar{A}) = \frac{\mathrm{Var}(\bar{A}^+) + \mathrm{Var}(\bar{A}^-) + 2\mathrm{Cov}(\bar{A}^+, \bar{A}^-)}{4} = \frac{\mathrm{Var}(A^+_m) + \mathrm{Cov}(A^+_m, A^-_m)}{2n},$$

where the n in the denominator comes from the pairwise independence of the $(A^+_{m,i}, A^-_{m,i})$, $i = 1, 2, \ldots, n$. If we are lucky enough to have $\mathrm{Cov}(A^+_m, A^-_m) \leq 0$, which is certainly intuitively pleasing, then the door opens for the antithetic estimator $\bar{A}$ to achieve a substantial variance reduction compared to that of the original naive estimator $\bar{A}^+$ (Glasserman [8]).

Example 1 We compare the performance of the naive and antithetic estimators $\bar{A}^+$ and $\bar{A}$ for $\mathrm{E}[A^+_m]$. To undertake this evaluation, we conducted a battery of MC simulation runs for $s_0 = 1$ and various values of m, ϕ, and τ^2. Illustrative MC results are given in Table 1, where for each parameter setting, we ran $n = 10{,}000$ independent replications of A^+_m and A^-_m to obtain the naive and antithetic estimators, $\bar{A}^+$ and $\bar{A}$. Recall that both estimators are unbiased for $\mathrm{E}[A^+_m]$; so we are interested in comparing $\mathrm{Var}[\bar{A}^+]$ and $\mathrm{Var}[\bar{A}]$, or almost equivalently, their respective standard errors, s.e.$[\bar{A}^+]$ and s.e.$[\bar{A}]$, where s.e.$[\bar{A}]$ is defined analogously to s.e.$[\bar{A}^+]$ in Equation (10). As the table clearly shows, s.e.$[\bar{A}^+]$/s.e.$[\bar{A}] \geq 3.07$ for the 8 cases depicted in Table 1, which indicates that $\bar{A}$ brings about at least a 9-fold variance reduction for those cases. All of this is achieved for only a minor increase in computational effort. □

This good performance of the antithetic estimator $\bar{A}$ of $\mathrm{E}[\bar{A}^+_m]$ is actually well known in the literature (Glasserman [8]). And in fact, we were already able to derive

Table 1: Standard errors ($\times 10^3$) of $\bar{A}^+$ and $\bar{A}$ as estimators of $\mathrm{E}[A_m^+]$ based on $n =$ 10,000 independent replications. For all cases, $s_0 = 1$.

m	ϕ	τ^2	s.e. $\bar{A}^+$	s.e. $\bar{A}$
4	0.01	0.03	3.46	0.81
		0.05	3.52	0.86
	0.02	0.03	3.62	0.87
		0.05	3.58	0.85
8	0.01	0.03	5.14	1.57
		0.05	5.04	1.62
	0.02	0.03	5.49	1.79
		0.05	5.35	1.73

an exact expression for $\mathrm{E}[A_m^+]$ via our Equation (9). Our task is now to provide more-general, unbiased, low-variance estimators. To this end, in §2.3.3, we will derive exact expressions for the variances of those more-general, unbiased estimators for $\mathrm{E}[A_m^+]$. These estimators are based on quasi-MC simulation, of which antithetics are a very special case.

2.3.3 Exact Results for More-General Estimators

Equation (8) gives the arithmetic average via the expression A_m^+ which incorporates a "+" sign in front of the i.i.d. normal increments $X_1, X_2, \ldots, X_m$, while Equation (11) does the same via the expression A_m^- having a "$-$" sign in front of the X_i's. In §2.3.2, we exploited the (hopefully) negative covariance between A_m^+ and A_m^- to obtain the antithetic estimator $\bar{A}$, which is unbiased for $\mathrm{E}[A_m^+]$ while having lower variance than the naive estimator that is based solely on realizations of the form A_m^+. In what follows, we will come up with a more-general covariance expression incorporating increments of the arbitrary form $\pm X_1, \pm X_2, \ldots, \pm X_m$. This expression will then suggest an estimator for $\mathrm{E}[A_m^+]$ that is even more efficient than $\bar{A}$.

For this purpose, consider the m-vector $\boldsymbol{X}$, whose components $X_1, X_2, \ldots, X_m$ are i.i.d. $\mathrm{Nor}(0, \tau^2)$. Let $\boldsymbol{a}$ and $\boldsymbol{b}$ denote m-vectors consisting of ± 1's, of which there are of course 2^m possible choices. In particular, $\boldsymbol{a} = (1,1,\ldots,1)$ corresponds to the estimator A_m^+ (based on one replication of $\boldsymbol{X}$), while $\boldsymbol{b} = (-1,-1,\ldots,-1)$ corresponds to the estimator A_m^- (based on the same single replication of $\boldsymbol{X}$). At this point, let

$$H_i(\boldsymbol{a}) \;\equiv\; \phi i + \sum_{\ell=1}^{i} a_\ell X_\ell \;\sim\; \mathrm{Nor}(\phi i, \tau^2 i), \quad i = 1,2,\ldots,m,$$

$A \equiv \frac{s_0}{m}\sum_{i=1}^m e^{H_i(\boldsymbol{a})}$, and similarly, $B \equiv \frac{s_0}{m}\sum_{i=1}^m e^{H_i(\boldsymbol{b})}$. The quantities A and B are expressions for arithmetic equity averages and have the same marginal distribution as A_m^+. Thus, by the arguments leading Equation (9), we have $\mathrm{E}[A] = \mathrm{E}[B] = f_m(\theta)$.

The goal now is to calculate $\mathrm{Cov}(A,B)$. To do so, suppose that $i \le j$. Then for any two m-vectors of ± 1's $\boldsymbol{a}$ and $\boldsymbol{b}$, we have

$$\begin{aligned} H_i(\boldsymbol{a}) + H_j(\boldsymbol{b}) &= \phi(i+j) + \sum_{\ell=1}^{i}(a_\ell + b_\ell)X_\ell + \sum_{\ell=i+1}^{j} b_\ell X_\ell \\ &\sim \mathrm{Nor}\left(\phi(i+j),\, \sum_{\ell=1}^{i}(a_\ell+b_\ell)^2\tau^2 + \sum_{\ell=i+1}^{j} b_\ell^2\tau^2\right) \\ &\sim \mathrm{Nor}\left(\phi(i+j),\, \left(2i + 2\sum_{\ell=1}^{i} a_\ell b_\ell + j - i\right)\tau^2\right). \end{aligned}$$

So, in general,

$$H_i(\boldsymbol{a}) + H_j(\boldsymbol{b}) \;\sim\; \mathrm{Nor}\big(\phi(i+j), (i+j+2D_{i\wedge j})\tau^2\big), \tag{12}$$

where

$$D_i \;\equiv\; D_i(\boldsymbol{a},\boldsymbol{b}) \;\equiv\; \sum_{\ell=1}^{i} a_\ell b_\ell, \quad i = 1,2,\ldots,m,$$

and $i \wedge j \equiv \min(i,j)$.

With no loss of generality, we henceforth take $s_0 = 1$. Then Equation (12) and properties of the lognormal distribution imply

$$
\begin{aligned}
m^2\mathrm{E}[AB] &= \sum_{i=1}^{m}\sum_{j=1}^{m}\mathrm{E}\big[e^{H_i(\boldsymbol{a})+H_i(\boldsymbol{b})}\big] \\
&= \sum_{i=1}^{m}\sum_{j=1}^{m} e^{(\phi+\frac{\tau^2}{2})(i+j)+D_{i\wedge j}\tau^2} \\
&= \sum_{i=1}^{m}\sum_{j=1}^{m} e^{\theta(i+j)+D_{i\wedge j}\tau^2} \\
&= \sum_{i=1}^{m} e^{2\theta i+D_i\tau^2} + 2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} e^{\theta(i+j)+D_i\tau^2} \\
&= \sum_{i=1}^{m} e^{2\theta i+D_i\tau^2} + 2\sum_{i=1}^{m-1} e^{\theta i+D_i\tau^2}e^{\theta(i+1)}\sum_{j=0}^{m-(i+1)} e^{\theta j} \\
&= \sum_{i=1}^{m} e^{2\theta i+D_i\tau^2} + 2e^{\theta}\sum_{i=1}^{m} e^{2\theta i+D_i\tau^2}\frac{(1-e^{\theta(m-i)})}{1-e^{\theta}} \\
&= \sum_{i=1}^{m}\left(1+\frac{2e^{\theta}}{1-e^{\theta}}\right)e^{2\theta i+D_i\tau^2} - \frac{2e^{\theta}}{1-e^{\theta}}\sum_{i=1}^{m} e^{2\theta i+D_i\tau^2+\theta(m-i)} \\
&= \sum_{i=1}^{m}\left\{\left(\frac{1+e^{\theta}}{1-e^{\theta}}\right)e^{2\theta i+D_i\tau^2} - \frac{2e^{\theta(m+1)}}{1-e^{\theta}}e^{\theta i+D_i\tau^2}\right\}, \qquad (13)
\end{aligned}
$$

from which we can readily calculate $\mathrm{Cov}(A,B) = \mathrm{E}[AB] - \mathrm{E}[A]\mathrm{E}[B]$.

Let $\alpha \equiv 2\theta+\tau^2$, $\beta \equiv 2\theta-\tau^2$, $\gamma \equiv \theta+\tau^2$, and $\delta \equiv \theta-\tau^2$. We will first calculate $\mathrm{Var}(A)$ by setting $\boldsymbol{a}=\boldsymbol{b}$ (so that $D_i = i$ for all i) and applying Equation (13); then

$$
\begin{aligned}
\mathrm{Var}(A) \;=\; \mathrm{Var}(B) &= \frac{1}{m^2}\sum_{i=1}^{m}\left\{\left(\frac{1+e^{\theta}}{1-e^{\theta}}\right)e^{\alpha i} - \frac{2e^{\theta(m+1)}}{1-e^{\theta}}e^{\gamma i}\right\} - f_m^2(\theta) \\
&= \left(\frac{1+e^{\theta}}{1-e^{\theta}}\right)\frac{f_m(\alpha)}{m} - \frac{2e^{\theta(m+1)}}{1-e^{\theta}}\frac{f_m(\gamma)}{m} - f_m^2(\theta). \qquad (14)
\end{aligned}
$$

Notice that $\mathrm{Var}(A)$ is independent of the choice of the ± 1 vector $\boldsymbol{a}$. On the other hand, $\mathrm{Cov}(A,B)$ depends on $\boldsymbol{a}$ and $\boldsymbol{b}$. In particular, consider any antithetic choice for which $D_i = -i$ for all i. For example, for $m=4$, we might choose

$$\boldsymbol{a} \;=\; (1,1,1,1) \quad \text{and} \quad \boldsymbol{b} \;=\; (-1,-1,-1,-1)$$

or

$$\boldsymbol{a} \;=\; (1,-1,1,-1) \quad \text{and} \quad \boldsymbol{b} \;=\; (-1,1,-1,1),$$

both of which result in $D_i = \sum_{\ell=1}^{i} a_\ell b_\ell = -i$ for $i = 1, 2, 3, 4$.

After some tedious but straightforward algebra similar to that leading to (14), we have

$$\mathrm{Cov}(A, B) \;=\; \left(\frac{1+e^\theta}{1-e^\theta}\right)\frac{f_m(\beta)}{m} - \frac{2e^{\theta(m+1)}}{1-e^\theta}\frac{f_m(\delta)}{m} - f_m^2(\theta). \tag{15}$$

Then Equations (14) and (15) imply

$$\begin{aligned}\mathrm{Var}\left(\frac{A+B}{2}\right) &= \frac{\mathrm{Var}(A) + \mathrm{Var}(B) + 2\mathrm{Cov}(A, B)}{4} = \frac{\mathrm{Var}(A) + \mathrm{Cov}(A, B)}{2} \\ &= \left(\frac{1+e^\theta}{1-e^\theta}\right)\frac{[f_m(\alpha) + f_m(\beta)]}{2m} - \frac{e^{\theta(m+1)}}{1-e^\theta}\frac{[f_m(\gamma) + f_m(\delta)]}{m} - f_m^2(\theta).\end{aligned} \tag{16}$$

It is this quantity that we hope is less than $\mathrm{Var}(A)$, in which case we will have a variance reduction. For instance, the very special case in which $\boldsymbol{a} = (1, 1, \ldots, 1)$ and $\boldsymbol{b} = (-1, -1, \ldots, -1)$ is treated in Example 1, where A and B in Equations (14) and (16) correspond to A_m^+ and A_m^-, respectively; thus, Equation (14) gives an exact expression for $\mathrm{Var}(A_m^+) = \mathrm{Var}(A_m^-)$ and Equation (16) gives an exact expression for $\mathrm{Var}(\frac{A_m^+ + A_m^-}{2}) = \mathrm{Var}(\bar{A}_m)$.

Figures 1 and 2 illustrate comparisons between the exact standard deviations of the naive estimator A_m^+ and the antithetic estimator $\bar{A}_m$ (both of which are based on a single replication of the underlying increments $\boldsymbol{X}$). The standard deviation values are calculated from Equations (14) and (16) for $s_0 = 1$ and various ϕ and τ^2 for the cases $m = 8$ and $m = 32$. Similar to Example 1, which employed Monte Carlo sampling (n replications of $\boldsymbol{X}$), our exact results in the figures again illustrate the potential variance reduction benefits that arise from the use of antithetics.

2.3.4 Full Quasi-Monte Carlo Estimator for $\mathrm{E}[\boldsymbol{A_m^+}]$

Since the antithetic average $\bar{A}_m$ works well as an estimator of $\mathrm{E}[A_m^+]$, we now see what can be gained by considering averages based on more-general combinations of

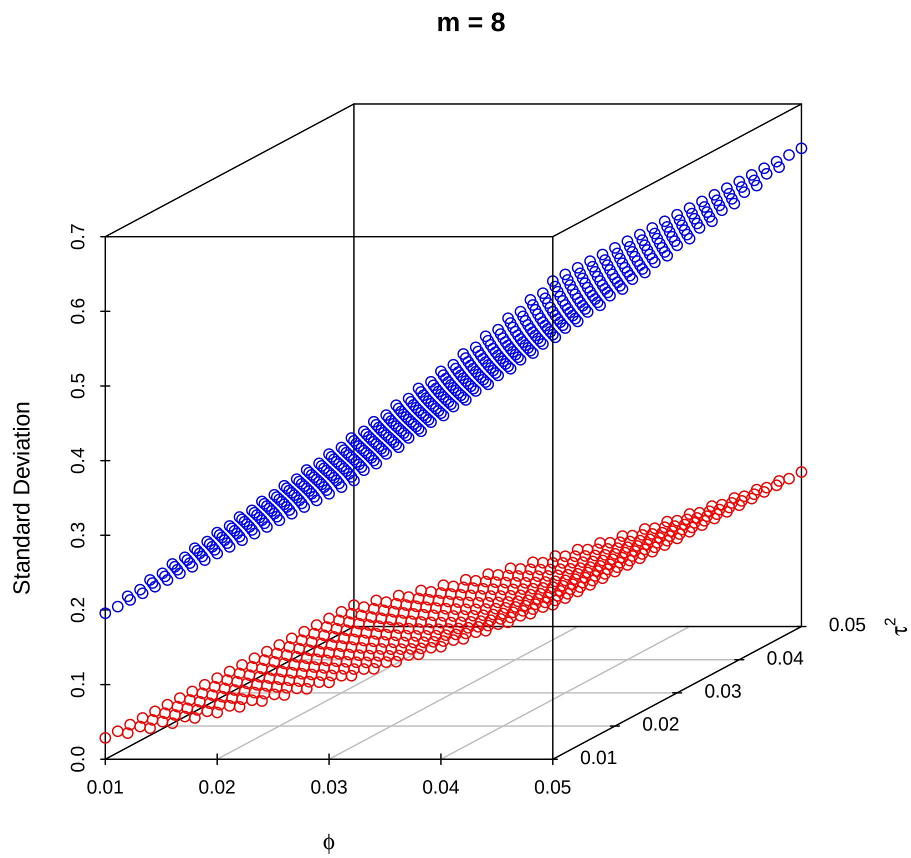

Figure 1: Exact standard deviations of the naive estimator A_m^+ from Equation (14) (in blue) and the antithetic estimator $\bar{A}_m$ from Equation (16) (in red) for $s_0 = 1$, various ϕ and τ^2, and $m = 8$.

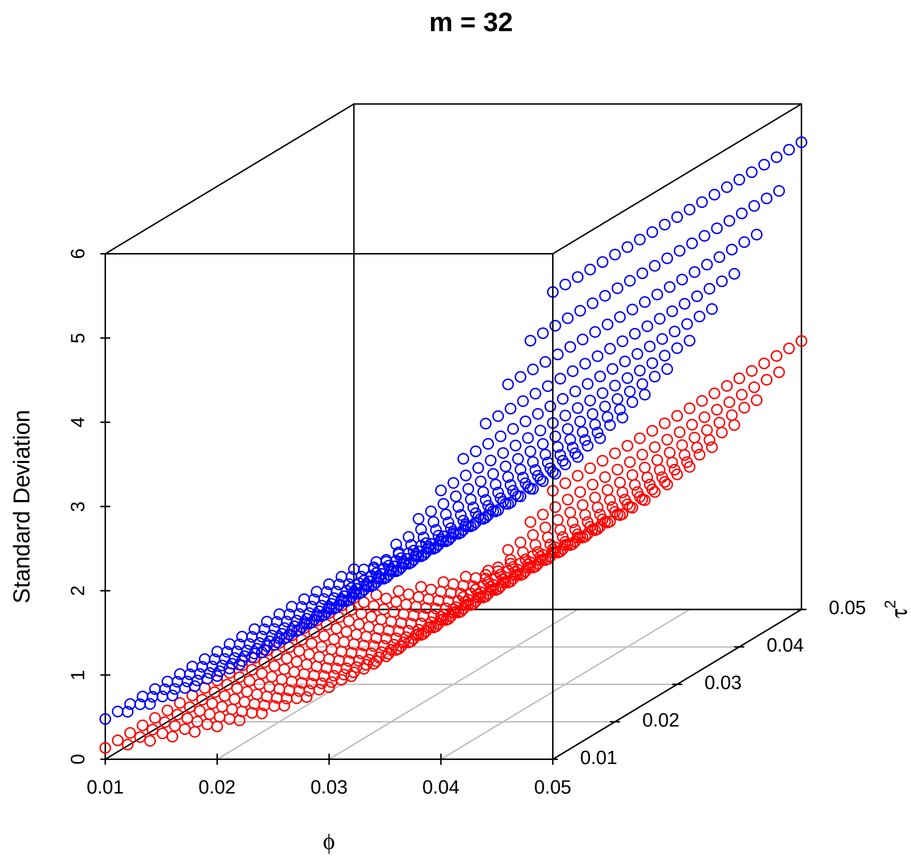

Figure 2: Exact standard deviations of the naive estimator A_m^+ from Equation (14) (in blue) and the antithetic estimator $\bar{A}_m$ from Equation (16) (in red) for $s_0 = 1$, various ϕ and τ^2, and $m = 32$.

± 1-vectors $\boldsymbol{a}$. We now consider all of the possible 2^m combinations of ± 1's. To this end, let the notation $[\boldsymbol{a}]^{j,m}$ denote the m-vector corresponding to the jth such combination, $j = 1, 2, \ldots, 2^m$, where we (arbitrarily) order the vectors by giving 1's precedence over -1's. For example, for the case $m = 4$, there are $2^m = 16$ possible vectors, namely,

$$[\boldsymbol{a}]^{1,4} = (1,1,1,1),\ [\boldsymbol{a}]^{2,4} = (1,1,1,-1),\ \ldots,\ [\boldsymbol{a}]^{16,4} = (-1,-1,-1,-1).$$

With the purpose of achieving more-substantial variance reductions in mind, we are interested in the grand average taken over Asian averages $\widetilde{A}_m^{(j)}$ formed from all of the 2^m possible ± 1-combinations,

$$\widetilde{A}_m \equiv \frac{1}{2^m} \sum_{j=1}^{2^m} \widetilde{A}_m^{(j)}, \tag{17}$$

where

$$\widetilde{A}_m^{(j)} \equiv \frac{1}{m} \sum_{i=1}^{m} e^{H_i([\boldsymbol{a}]^{j,m})} = \frac{1}{m} \sum_{i=1}^{m} \exp\left\{ \phi i + \sum_{\ell=1}^{i} [a]_\ell^{j,m} X_\ell \right\}, \quad j = 1, 2, \ldots, 2^m. \tag{18}$$

In order to simplify the notational fog, for $\ell = 1, 2, \ldots, m$, let $Y_\ell^+ \equiv \exp\{\phi + X_\ell\}$, $Y_\ell^- \equiv \exp\{\phi - X_\ell\}$, and

$$Y_\ell([\boldsymbol{a}]^{j,m}) \equiv \begin{cases} Y_\ell^+ & \text{if } [a]_\ell^{j,m} = 1 \\ Y_\ell^- & \text{if } [a]_\ell^{j,m} = -1. \end{cases}$$

Then Equations (17)–(18) become

$$\widetilde{A}_m = \frac{1}{m2^m} \sum_{j=1}^{2^m} \sum_{i=1}^{m} \prod_{\ell=1}^{i} Y_\ell([\boldsymbol{a}]^{j,m}). \tag{19}$$

For instance, for the case $m = 3$, Equation (19) becomes

$$\begin{aligned} \widetilde{A}_3 = \frac{1}{24} \Big\{ & \left[Y_1^+ + Y_1^+ Y_2^+ + Y_1^+ Y_2^+ Y_3^+ \right] + \left[Y_1^+ + Y_1^+ Y_2^+ + Y_1^+ Y_2^+ Y_3^- \right] \\ & + \cdots + \left[Y_1^- + Y_1^- Y_2^- + Y_1^- Y_2^- Y_3^- \right] \Big\}. \end{aligned}$$

In addition, for $k = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, 2^{m-k}$, let $V_{k,j}$ denote the jth "Asian sum" arising from observations $X_k, X_{k+1}, \ldots, X_m$. i.e.,

$$V_{k,j} \equiv \sum_{i=1}^{m-k+1} \prod_{\ell=1}^{i} Y_\ell([\boldsymbol{a}]^{j,m-k+1}). \tag{20}$$

Thus, for example, for the case $k = 1$ and $m = 4$, we have

$$\begin{aligned}
V_{1,1} &= Y_1^+ + Y_1^+ Y_2^+ + Y_1^+ Y_2^+ Y_3^+ + Y_1^+ Y_2^+ Y_3^+ Y_4^+ \\
V_{1,2} &= Y_1^+ + Y_1^+ Y_2^+ + Y_1^+ Y_2^+ Y_3^+ + Y_1^+ Y_2^+ Y_3^+ Y_4^- \\
&\vdots \\
V_{1,16} &= Y_1^- + Y_1^- Y_2^- + Y_1^- Y_2^- Y_3^- + Y_1^- Y_2^- Y_3^- Y_4^- .
\end{aligned}$$

For the case $k = 2$ and $m = 4$, which only involves observations X_2, X_3, X_4, we have

$$\begin{aligned}
V_{2,1} &= Y_2^+ + Y_2^+ Y_3^+ + Y_2^+ Y_3^+ Y_4^+ \\
V_{2,2} &= Y_2^+ + Y_2^+ Y_3^+ + Y_2^+ Y_3^+ Y_4^- \\
&\vdots \\
V_{2,8} &= Y_2^- + Y_2^- Y_3^- + Y_2^- Y_3^- Y_4^- .
\end{aligned}$$

We finally define $\widetilde{Y}_i \equiv Y_i^+ + Y_i^-$ for $i = 1, 2, \ldots, m$, and

$$\widetilde{V}_k \equiv \sum_{j=1}^{2^{m-k+1}} V_{k,j}, \quad k = 1, 2, \ldots, m. \tag{21}$$

Then we have

$$
\begin{aligned}
\widetilde{V}_1 &= \sum_{j=1}^{2^m} V_{1,j} \\
&= \Big\{ \big[Y_1^+ + Y_1^+ Y_2^+ + \cdots + (Y_1^+ Y_2^+ \cdots Y_m^+)\big] \\
&\quad + \big[Y_1^+ + Y_1^+ Y_2^+ + \cdots + (Y_1^+ Y_2^+ \cdots Y_m^-)\big] \\
&\quad + \cdots + \big[Y_1^+ + Y_1^- Y_2^- + \cdots + (Y_1^- Y_2^- \cdots Y_m^-)\big] \Big\} \\
&\quad + \Big\{ \big[Y_1^- + Y_1^+ Y_2^+ + \cdots + (Y_1^+ Y_2^+ \cdots Y_m^+)\big] \\
&\quad + \big[Y_1^- + Y_1^+ Y_2^+ + \cdots + (Y_1^+ Y_2^+ \cdots Y_m^-)\big] \\
&\quad + \cdots + \big[Y_1^- + Y_1^- Y_2^- + \cdots + (Y_1^- Y_2^- \cdots Y_m^-)\big] \Big\} \\
&= \widetilde{Y}_1 \Big\{ \big[1 + Y_2^+ + \cdots + (Y_2^+ Y_3^+ \cdots Y_m^+)\big] \\
&\quad + \big[1 + Y_2^+ + \cdots + (Y_2^+ Y_3^+ \cdots Y_m^-)\big] \\
&\quad + \cdots + \big[1 + Y_2^- + \cdots + (Y_2^- Y_3^- \cdots Y_m^-)\big] \Big\} \\
&= \widetilde{Y}_1 (2^{m-1} + \widetilde{V}_2) \quad \text{(by Equation (21))} \\
&= \widetilde{Y}_1 [2^{m-1} + \widetilde{Y}_2 (2^{m-2} + \widetilde{V}_3)] \\
&= 2^{m-1} \widetilde{Y}_1 + 2^{m-2} \widetilde{Y}_1 \widetilde{Y}_2 + \cdots + 2^1 (\widetilde{Y}_1 \widetilde{Y}_2 \cdots \widetilde{Y}_{m-1}) + (\widetilde{Y}_1 \widetilde{Y}_2 \cdots \widetilde{Y}_{m-1}) \widetilde{V}_m \\
&= \sum_{i=1}^{m} 2^{m-i} \prod_{j=1}^{i} \widetilde{Y}_j \quad (\text{since } \widetilde{V}_m = \widetilde{Y}_m). \qquad (22)
\end{aligned}
$$

Therefore, by Equations (19), (20) (with $k = 1$), and (22), we have

$$\widetilde{A}_m = \frac{\widetilde{V}_1}{m 2^m} = \frac{1}{m} \sum_{i=1}^{m} 2^{-i} \prod_{j=1}^{i} \widetilde{Y}_j, \qquad (23)$$

which is very easy to calculate; cf. the methodology presented in Carverhill and Clewlow [7]. Now suppose that $\widetilde{A}_{m,1}, \widetilde{A}_{m,2}, \ldots, \widetilde{A}_{m,n}$ are n i.i.d. replications of $\widetilde{A}_m$ (based on n i.i.d. replications of $\boldsymbol{X}$). We refer to $\widetilde{A} \equiv \frac{1}{n} \sum_{i=1}^{n} \widetilde{A}_{m,i}$ as the *full quasi-Monte Carlo* (FQMC) estimator for $\mathrm{E}[A_m^+]$; by construction, it is unbiased for $\mathrm{E}[A_m^+]$. The goal is of course to achieve substantial variance reductions compared to other estimators for $\mathrm{E}[A_m^+]$. We could ostensibly use Equations (23) and (16) along with some extremely tedious algebra to calculate $\mathrm{Var}(\widetilde{A})$; but for now, we will resort to MC simulation to evaluate this quantity.

Example 2 We repeat Example 1, but here we add in results for the new estimator. Thus, we compare the performance of the naive, antithetic, and FQMC estimators $\bar{A}^+$, $\bar{A}$, and $\widetilde{A}$ for $\mathrm{E}[A_m^+]$. As before, the results are based on $n = 10{,}000$ independent replications for each estimator for the cases $s_0 = 1$ and various values of m, ϕ, and τ^2. The results are given in Table 2, which provides the standard errors of the three estimators for each parameter setting. Roughly speaking, the use of the FQMC estimator $\widetilde{A}$ achieves a 50% reduction in the standard error (or a 75% reduction in the variance) compared to the antithetic estimator $\bar{A}$. $\square$

Table 2: Standard errors ($\times 10^3$) of $\bar{A}^+$, $\bar{A}$, and $\widetilde{A}$ as estimators of $\mathrm{E}[A_m^+]$ based on $n = 10{,}000$ independent replications. For all cases, $s_0 = 1$.

m	ϕ	τ^2	s.e.		
			$\bar{A}^+$	$\bar{A}$	$\widetilde{A}$
4	0.01	0.03	3.46	0.81	0.53
		0.05	3.52	0.86	0.54
	0.02	0.03	3.62	0.87	0.56
		0.05	3.58	0.85	0.55
8	0.01	0.03	5.14	1.57	0.76
		0.05	5.04	1.62	0.77
	0.02	0.03	5.49	1.79	0.81
		0.05	5.35	1.73	0.80

We will also have the need in the sequel for $\widetilde{V}_1, \widetilde{V}_2, \ldots, \widetilde{V}_m$. To this end, it is easy to see from Equation (22) (but now based on observations $X_k, X_{k+1}, \ldots, X_m$) that

$$\widetilde{V}_k = \sum_{i=1}^{m-k+1} 2^{m-k+1-i} \prod_{j=k}^{i+k-1} \widetilde{Y}_j, \quad k = 1, 2, \ldots, m. \tag{24}$$

We calculate $\widetilde{V}_m, \widetilde{V}_{m-1}, \ldots, \widetilde{V}_1$ by working backwards.

$$
\begin{aligned}
\widetilde{V}_m &= \widetilde{Y}_m \\
\widetilde{V}_{m-1} &= 2\widetilde{Y}_{m-1} + \widetilde{Y}_{m-1}\widetilde{Y}_m \;=\; \widetilde{Y}_{m-1}(2 + \widetilde{V}_m) \\
\widetilde{V}_{m-2} &= 4\widetilde{Y}_{m-2} + 2\widetilde{Y}_{m-2}\widetilde{Y}_{m-1} + \widetilde{Y}_{m-2}\widetilde{Y}_{m-1} \;=\; \widetilde{Y}_{m-2}(2^2 + \widetilde{V}_{m-1}) \\
&\vdots \\
\widetilde{V}_{m-k} &= \widetilde{Y}_{m-k}(2^k + \widetilde{V}_{m-k+1}), \quad k = 1, 2, \ldots, m-1. \qquad (25)
\end{aligned}
$$

2.3.5 Higher-Order Moments

We can develop analogous technology to estimate higher-order moments $\mathrm{E}[(A_m^+)^\ell]$. In order to do so, define $\widetilde{Y}_i^{(\ell)} \equiv (Y_i^+)^\ell + (Y_i^-)^\ell$ for $i = 1, 2, \ldots, m$ and $\ell = 1, 2, \ldots$, and

$$
\widetilde{V}_k^{(\ell)} \equiv \sum_{j=1}^{2^{m-i+1}} V_{k,j}^\ell, \quad \text{for } k = 1, 2, \ldots, m \text{ and } \ell = 1, 2, \ldots. \qquad (26)
$$

In particular,

$$
\begin{aligned}
\widetilde{V}_1^{(\ell)} &= \sum_{j=1}^{2^m} V_{1,j}^{\ell} \\
&= \Big\{ \big[Y_1^+ + Y_1^+Y_2^+ + \cdots + (Y_1^+Y_2^+\cdots Y_m^+)\big]^{\ell} \\
&\quad + \big[Y_1^+ + Y_1^+Y_2^+ + \cdots + (Y_1^+Y_2^+\cdots Y_m^-)\big]^{\ell} \\
&\quad + \cdots + \big[Y_1^+ + Y_1^-Y_2^- + \cdots + (Y_1^-Y_2^-\cdots Y_m^-)\big]^{\ell} \Big\} \\
&\quad + \Big\{ \big[Y_1^- + Y_1^+Y_2^+ + \cdots + (Y_1^+Y_2^+\cdots Y_m^+)\big]^{\ell} \\
&\quad + \big[Y_1^- + Y_1^+Y_2^+ + \cdots + (Y_1^+Y_2^+\cdots Y_m^-)\big]^{\ell} \\
&\quad + \cdots + \big[Y_1^- + Y_1^-Y_2^- + \cdots + (Y_1^-Y_2^-\cdots Y_m^-)\big]^{\ell} \Big\} \\
&= \widetilde{Y}_1^{(\ell)} \Bigg\{ \Bigg[1 + \Bigg(Y_2^+ + \cdots + (Y_2^+Y_3^+\cdots Y_m^+)\Bigg)\Bigg]^{\ell} \\
&\quad + \Bigg[1 + \Bigg(Y_2^+ + \cdots + (Y_2^+Y_3^+\cdots Y_m^-)\Bigg)\Bigg]^{\ell} \\
&\quad + \cdots + \Bigg[1 + \Bigg(Y_2^- + \cdots + (Y_2^-Y_3^-\cdots Y_m^-)\Bigg)\Bigg]^{\ell} \Bigg\} \\
&= \widetilde{Y}_1^{(\ell)} \left[2^{m-1} + \sum_{p=1}^{\ell-1} \binom{\ell}{p} \widetilde{V}_2^{(p)} + \widetilde{V}_2^{(\ell)}\right] \\
&= \widetilde{Y}_1^{(\ell)} \left\{2^{m-1} + \sum_{p=1}^{\ell-1} \binom{\ell}{p} \widetilde{V}_2^{(p)} + \left[\widetilde{Y}_2^{(\ell)}\left(2^{m-2} + \sum_{p=1}^{\ell-1} \binom{\ell}{p} \widetilde{V}_3^{(p)} + \widetilde{V}_3^{(\ell)}\right)\right]\right\} \\
&= \widetilde{Y}_1^{(\ell)} \left[2^{m-1} + \sum_{p=1}^{\ell-1} \binom{\ell}{p} \widetilde{V}_2^{(p)}\right] + \widetilde{Y}_1^{(\ell)}\widetilde{Y}_2^{(\ell)} \left[2^{m-2} + \sum_{p=1}^{\ell-1} \binom{\ell}{p} \widetilde{V}_3^{(p)}\right] + \widetilde{Y}_1^{(\ell)}\widetilde{Y}_2^{(\ell)}\widetilde{V}_3^{(\ell)} \\
&= \sum_{i=1}^{m-1} \left\{ \left(\prod_{j=1}^{i} \widetilde{Y}_j^{(\ell)}\right) \left[2^{m-i} + \sum_{p=1}^{\ell-1} \binom{\ell}{p} \widetilde{V}_{i+1}^{(p)}\right] \right\} + \left(\widetilde{Y}_1^{(\ell)}\widetilde{Y}_2^{(\ell)}\cdots\widetilde{Y}_{m-1}^{(\ell)}\right)\widetilde{V}_m^{(\ell)} \\
&= \sum_{i=1}^{m} \left(\prod_{j=1}^{i} \widetilde{Y}_j^{(\ell)}\right) \left[2^{m-i} + \sum_{p=1}^{\ell-1} \binom{\ell}{p} \widetilde{V}_{i+1}^{(p)}\right] \quad \text{(since } \widetilde{V}_{i+1}^{(p)} = 0 \text{ for all } i\text{)}. \qquad (27)
\end{aligned}
$$

Similarly, for $k = 1, 2, \ldots, m$,

$$
\widetilde{V}_k^{(\ell)} = \sum_{i=1}^{m-k+1} \left(\prod_{j=k}^{i+k-1} \widetilde{Y}_j^{(\ell)}\right) \left[2^{m-i-k+1} + \sum_{p=1}^{\ell-1} \binom{\ell}{p} \widetilde{V}_{i+k}^{(p)}\right]. \qquad (28)
$$

Using the same reasoning as that leading to Equation (25), we can calculate $\widetilde{V}_m^{(\ell)}, \widetilde{V}_{m-1}^{(\ell)}, \ldots, \widetilde{V}_1^{(\ell)}$ by working backwards. The result is that for $k = 1, 2, \ldots, m-1$

and $\ell = 1, 2, \ldots$, we have

$$\widetilde{V}_{m-k}^{(\ell)} \;=\; \widetilde{Y}_{m-k}^{(\ell)}\left[2^k + \sum_{p=1}^{\ell} \binom{\ell}{p} \widetilde{V}_{m-k+1}^{(p)}\right]. \tag{29}$$

The following is a summary algorithm to calculate all of the $\widetilde{V}_i^{(\ell)}$'s.

Algorithm

For $\ell = 1, 2, \ldots, 6$

  Set $\widetilde{V}_m^{(\ell)} \leftarrow \widetilde{Y}_m^{(\ell)}$

  For $k = 1, 2, \ldots, m-1$

    Set $\widetilde{V}_{m-k}^{(\ell)} \leftarrow$ (29)

  next k

next ℓ

2.3.6 Game Plan for Arithmetic Asian Options

All of our work so far in §2.3 has involved the estimation of moments of Asian sums. But what about the analogous options on those sums? Recall that the Asian call option associated with $A_m \equiv A_m^+$ at expiry time T for strike price k is given by $C_{A_m} \equiv (A_m - k)^+$, where $(y)^+ = \max\{y, 0\}$. The call's fair (expected) value at expiry, discounted back to time 0, is

$$c_{A_m} \;\equiv\; e^{-rT}\mathrm{E}[(A_m - k)^+] \;=\; e^{-rT}\mathrm{E}[(\widetilde{A}_{j,m} - k)^+], \quad j = 1, 2, \ldots, 2^m.$$

Estimation of the quantity c_{A_m} is somewhat more-difficult than that of $\mathrm{E}[A_m]$ owing to the tricky $(\cdot)^+$ term, which does not come out in closed form. There are a variety of methods in the literature to attempt the task — everything from series approximations to characteristic function methods to Monte Carlo techniques. We will build on the MC work of the previous subsection, though we will not be able to use a single tidy estimator such as $\widetilde{A}$ for the job. Nevertheless, we will take advantage of the fact that a single realization of m i.i.d. normal increments $\boldsymbol{X} \equiv (X_1, X_2, \ldots, X_m)$

can be used to generate the 2^m individual quasi-Monte Carlo $\widetilde{A}_m^{(j)}$, $j = 1, 2, \ldots, 2^m$, corresponding to the various choices of ± 1 coefficients in front of those increments.

Thus, from the single realization $\boldsymbol{X}$, we could estimate c_{A_m} by, for example, the naive estimator,

$$\bar{C}_m^+ \;\equiv\; e^{-rT}(A_m^+ - k)^+, \tag{30}$$

the antithetic estimator,

$$\bar{C}_m \;\equiv\; \frac{e^{-rT}}{2}\big[(A_m^+ - k)^+ + (A_m^- - k)^+\big], \tag{31}$$

or the full quasi-Monte Carlo estimator,

$$\widetilde{C}_m \;\equiv\; \frac{e^{-rT}}{2^m}\sum_{j=1}^{2^m}(\widetilde{A}_m^{(j)} - k)^+. \tag{32}$$

Whatever the choice, there is a trade-off. The naive estimator $\bar{C}_m^+$ is trivial to calculate, but will have high variance; the antithetic estimator $\bar{C}_m$ is also easy to calculate, and will likely have reduced variance; and the FQMC estimator $\widetilde{C}_m$ will likely have even lower variance, but at the cost of potential burdensome calculation effort, especially as m increases — even though we only have to generate m i.i.d. normals, we do not currently have an efficient way of calculating the 2^m terms in the associated summation, which quickly becomes prohibitive. This issue will be discussed in §2.4.1 and then more so in Chapter 3.

Example 3 Parallel to Example 2, we compare the performance of the naive, antithetic, and FQMC estimators $\bar{C}^+$, $\bar{C}$, and $\widetilde{C}$ for c_{A_m}, where each is the sample mean of n = 10,000 independent replications of $\bar{C}_m^+$, $\bar{C}_m$, and $\widetilde{C}_m$, respectively. We take $s_0 = k = T = 1$ and consider various values of m, ϕ, and τ^2. The results are given in Table 3 for the undiscounted case, which provides the standard errors of the three estimators for each parameter setting. Roughly speaking, the use of the antithetic and FQMC estimators $\bar{C}$ and $\widetilde{C}$ achieves variance reductions over the naive estimator $\bar{C}^+$ that are comparable to those from Table 2. $\square$

Table 3: Standard errors ($\times 10^3$) of $\bar{C}^+$, $\bar{C}$, and $\widetilde{C}$ as estimators of the undiscounted option price $\mathrm{E}[(A_m^+ - k)^+]$ based on n = 10,000 independent replications. For all cases, $s_0 = k = T = 1$.

m	ϕ	τ^2	s.e. $\bar{C}^+$	s.e. $\bar{C}$	s.e. $\widetilde{C}$
4	0.01	0.03	2.68	1.42	0.89
		0.05	2.72	1.48	0.90
	0.02	0.03	2.89	1.49	0.93
		0.05	2.86	1.47	0.92
8	0.01	0.03	4.33	2.28	1.06
		0.05	4.21	2.31	1.08
	0.02	0.03	4.74	2.49	1.11
		0.05	4.60	2.43	1.11

2.4 Potpourri

This section is concerned with a variety of complementary topics on Asian averages. We start in §2.4.1 with a compromise estimator for $\mathrm{E}[A_m^+]$ and the associated arithmetic Asian option c_{A_m} that balances the efficiency / computation trade-off of the FQMC estimator. §2.4.2 discusses variance reductions that can be obtained by permuting the increments $X_1, X_2, \ldots, X_m$. §2.4.3 combines the quasi-Monte Carlo and permutation tricks to obtain additional variance reduction improvements. Finally, Section 2.4.4 introduces miscellany that will be addressed elsewhere in the thesis.

2.4.1 Partial Quasi-Monte Carlo Estimators for $\mathrm{E}[A_m^+]$ and c_{A_m}

We have already stated that the single-replication FQMC arithmetic Asian option estimator $\widetilde{C}_m$ has lower variance but requires greater computational effort than the analogous naive and antithetic estimators. This issue clearly arises due to the diminishing returns of incorporating more and more of the 2^m terms $\widetilde{A}_m^{(j)}$, $j = 1, 2, \ldots, 2^m$, in the FQMC.

To address this issue, we propose a compromise between the FQMC estimator and the traditional antithetic estimator. Namely, take the average of a certain subset of the $\widetilde{A}_m^{(j)}$'s with ± 1 coefficients that are likely to induce negative (or at least modest) correlations amongst the $\widetilde{A}_m^{(j)}$'s. We call the resulting estimator a *partial quasi-Monte Carlo* (PQMC) estimator.

There are many ways to undertake the task of selecting an appropriate subset for the PQMC estimator. Our method of choice is straightforward. With little loss of generality, suppose that $m = 2^\ell$ for some ℓ. Simply select all $\widetilde{A}_m^{(j)}$'s whose associated ± 1 vectors $\boldsymbol{a}$ consist of blocks of consecutive $+1$ coefficients and blocks of consecutive -1 coefficients all having lengths that are powers of two and at least $2^{\ell'}$, where ℓ' is a specified nonnegative integer $\leq \ell$. We illustrate via two simple examples.

Example 4 Suppose $m = 4$, so that $\ell = 2$. Note that there are $2^m = 16$ possible vectors $\boldsymbol{a}$ of ± 1's, as illustrated in Table 4. If we only consider coefficient vectors having all $+1$ and -1 blocks of at least size 2, so that $\ell' \geq 1$, then we can only use the first 4 vectors $\boldsymbol{a}$ of ± 1's from Table 4 (the remaining 12 possible vectors $\boldsymbol{a}$ contain at least one singleton $+1$ or -1). $\square$

Example 5 Suppose $m = 16$, so that $\ell = 4$. Table 5 lists 16 vectors $\boldsymbol{a}$ of ± 1's for which $\ell' \geq 1$. (There are other vectors for which $\ell' \geq 1$ that are not listed here, e.g., several vectors containing blocks having sizes that are not powers of 2.) We define $A_{1:j}$ as the PQMC estimator computed by averaging the j individual estimators arising from the $\boldsymbol{a}$'s of the first j rows of Table 5, $j = 1, 2, \ldots, 16$. Each $A_{1:j}$ is unbiased for $\mathrm{E}[A_m^+]$. Moreover, note that what we have called the naive estimator of $\mathrm{E}[A_m^+]$ is simply $A_m^+ = A_{1:1}$; and our antithetic estimator of $\mathrm{E}[A_m^+]$ is none other than $\bar{A}_m = A_{1:2}$.

While we might not expect the variance reduction using a PQMC estimator $A_{1:j}$ to be as dramatic as that using the FQMC estimator $\widetilde{A}_m$, we would still hope to achieve achieve a significant reduction in variance compared to A_m^+ and $\bar{A}_m$. And in any case, the calculation of $A_{1:j}$ (for moderate values of j) is much less costly than that of $\widetilde{A}_m$.

Table 4: Coefficient vectors $\boldsymbol{a}$ for use in the PQMC method for $m = 4$ with various minimum block sizes $2^{\ell'}$. A "+" denotes a coefficient of 1, and a "−" denotes a coefficient of −1. Rows are grouped by antithetic pairs.

ℓ'	a_1	a_2	a_3	a_4
2	+	+	+	+
	−	−	−	−
1	+	+	−	−
	−	−	+	+
0	+	+	+	−
	−	−	−	+
	+	+	−	+
	−	−	+	−
	+	−	+	+
	−	+	−	−
	+	−	+	−
	−	+	−	+
	+	−	−	+
	−	+	+	−
	+	−	−	−
	−	+	+	+

Table 6 reports MC results on estimator standard error performance for several PQMC estimators and the FQMC estimator. We ran $n = 100{,}000$ independent replications of each estimator for various ϕ and τ^2 values for the case $m = 16$ and $s_0 = 1$. The notations $\bar{A}^+, \bar{A}, \bar{A}_{1:4}, \ldots, \bar{A}_{1:16}, \widetilde{A}$ (previously defined or trivially obvious sample averages taken over the n replications) serve to indicate that the analogous column entries are the s.e.'s of the associated sample averages.

For fixed ϕ and τ^2, we see that the s.e.'s decrease quite rapidly for the first few $\bar{A}_{1:j}$'s and then slowly work their way down towards that of $\widetilde{A}$. Of course, it is easy to calculate $\widetilde{A}$ via Equation (23). But in light of the substantial computational burden to calculate the associated option value $\widetilde{C}_m$ from Equation (32), one may opt to use a less-expensive PQMC-based estimator. $\square$

Table 5: Coefficient vectors $\boldsymbol{a}$ for use in the PQMC method for $m = 16$ with various minimum block sizes $2^{\ell'}$. A "+" denotes a coefficient of 1, and a "−" denotes a coefficient of -1. Rows are grouped by antithetic pairs.

ℓ'	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
3	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+
2	+	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−
	−	−	−	−	+	+	+	+	−	−	−	−	+	+	+	+
	+	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
	−	−	−	−	+	+	+	+	+	+	+	+	−	−	−	−
1	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−
	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+
	+	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+
	−	−	+	+	−	−	+	+	+	+	−	−	+	+	−	−
	+	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
	−	−	+	+	+	+	−	−	−	−	+	+	+	+	−	−
	+	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−
	−	−	+	+	+	+	−	−	+	+	−	−	−	−	+	+

2.4.2 Permutations

An interesting characteristic about the Asian average is that the order in which the $X_1, X_2, \ldots, X_m$'s are taken matters. Roughly speaking, earlier observations are given more weight than subsequent observations — simply because the earlier observations are used in more terms of the Asian average. With this fact in mind, we can take a single realization $\boldsymbol{X}$ and turn it into $m!$ permuted realizations. If we let $\boldsymbol{X}_i$ denote the ith such permutation, $i = 1, 2, \ldots, m!$, we could calculate the arithmetic Asian averages arising from those permutations, say, $A_m(\boldsymbol{X}_i)$, $i = 1, 2, \ldots, m!$, each of which is unbiased for $\mathrm{E}[A_m^+]$. And then we could take the grand average of those permutations to obtain what we would hope is a low-variance estimator of $\mathrm{E}[A_m^+]$, namely,

$$\widetilde{A}_{P,m} \equiv \frac{1}{m!} \sum_{i=1}^{m!} A_m(\boldsymbol{X}_i).$$

Table 6: Standard errors ($\times 10^4$) of the PQMC estimators of $\mathrm{E}[A_m^+]$ using the coefficients displayed in Table 5 and based on n = 100,000 independent replications. For all cases, $s_0 = 1$ and $m = 16$.

ϕ	τ^2	s.e.									
		$\bar{A}^+$	$\bar{A}$	$\bar{A}_{1:4}$	$\bar{A}_{1:6}$	$\bar{A}_{1:8}$	$\bar{A}_{1:10}$	$\bar{A}_{1:12}$	$\bar{A}_{1:14}$	$\bar{A}_{1:16}$	$\widetilde{A}$
	0.04	3.96	0.40	0.34	0.28	0.26	0.23	0.21	0.20	0.19	0.14
0.04	0.12	6.93	1.21	1.04	0.84	0.79	0.68	0.64	0.60	0.58	0.42
	0.20	9.04	2.01	1.72	1.40	1.31	1.14	1.07	0.99	0.96	0.69
	0.04	4.09	0.42	0.36	0.29	0.27	0.24	0.22	0.21	0.20	0.14
0.10	0.12	7.17	1.24	1.07	0.87	0.81	0.71	0.67	0.62	0.60	0.43
	0.20	9.37	2.08	1.78	1.45	1.36	1.18	1.11	1.03	1.00	0.72
	0.04	4.25	0.43	0.37	0.30	0.28	0.25	0.23	0.21	0.21	0.15
0.16	0.12	7.49	1.33	1.12	0.91	0.85	0.74	0.70	0.64	0.62	0.45
	0.20	9.76	2.17	1.85	1.51	1.41	1.23	1.16	1.07	1.04	0.75

Since the $A_m(\boldsymbol{X}_i)$'s will likely have some degree of positive correlation, we might anticipate that $\widetilde{A}_{P,m}$ will achieve only a small-to-moderate reduction in variance compared to its predecessor estimators; and of course, the $m!$ terms in the summand come at a high computational cost if they are calculated one-by-one. Potential solution strategies include (i) developing an iterative procedure to calculate $A_m(\boldsymbol{X}_i)$ quickly and efficiently, and (ii) develop a compromise estimator that trades computation time and variance reduction in a spirit similar to that of the PQMC estimators of §2.4.1.

For now we discuss the latter compromise strategy (ii). In an attempt to minimize computation time yet achieve a nontrivial variance reduction, consider the average of the original Asian average using the realization $X_1, X_2, \ldots, X_m$ and a second Asian using the reversed realization $X_m, X_{m-1}, \ldots, X_1$,

$$\begin{aligned}\bar{A}_{P,m}^R &\equiv \frac{A_m(X_1, X_2, \ldots, X_m) + A_m(X_m, X_{m-1}, \ldots, X_1)}{2} \\ &= \frac{A_m^+ + A_m(X_m, X_{m-1}, \ldots, X_1)}{2}.\end{aligned}$$

Example 6 Yet again, we repeat our bellwether Example 1, but this time we add in results for the permutation estimators for $\mathrm{E}[A_m^+]$. Thus, we compare the performance of the naive, antithetic, FQMC, “reversed,” and fully permuted estimators, $\bar{A}^+$, $\bar{A}$, $\widetilde{A}$, $\bar{A}_P^R$, and $\widetilde{A}_P$, respectively. We ran $n = 10{,}000$ independent replications for each estimator for the cases $s_0 = 1$ and various values of m, ϕ, and τ^2. The m-values in this example were small enough so as to enable ready computation of the otherwise burdensome FQMC and permutation estimators.

The results are given in Table 7, which provides the standard errors of these five estimators (among others) for each parameter setting. Table 8 does the same for the analogous option pricing set-up, where we use the notations $\bar{C}^+$, $\bar{C}$, $\widetilde{C}$, $\bar{C}_P^R$, and $\widetilde{C}_P$ for the corresponding option estimators. Roughly speaking, we see that for the examples presented in the tables, the reversed and fully permuted estimators, $\bar{A}_P^R$ and $\widetilde{A}_P$, yield modest variance reductions as compared to the naive estimator, $\bar{A}^+$ (about 7–11% reductions in the standard error). However, the reversed and fully permuted estimators are much less efficient than the antithetic and FQMC estimators, $\bar{A}$ and $\widetilde{A}$. Moreover, the fully permuted estimator yields only a trivial variance improvement when compared to its reversed colleague — hardly worth all of the extra computation. §2.4.3 attempts to utilize these modest savings by combining the various methods discussed above. $\square$

2.4.3 Combining the Methods

Since all of the estimators studied so far are unbiased for their target parameters, we can combine them. In particular, one could ostensibly take the average of all of the $m!2^m$ FQMC and permutation estimators to obtain an unbiased grand estimator having variance that may be somewhat smaller than any of its constituents. Of course, one must be willing to overlook the daunting computation requirements, particularly for large values of m.

We define $\bar{\bar{A}}_1$ as the average of the 4 realizations obtained from $\boldsymbol{X}$ using the antithetic estimator together with the reversed permutation estimator. In addition,

Table 7: Standard errors ($\times 10^3$) of the naive, antithetic, FQMC, permutation, and combination estimators of $\mathrm{E}[A_m^+]$ based on $n = 10{,}000$ independent replications. For all cases, $s_0 = 1$.

m	ϕ	τ^2	s.e.						
			$\bar{A}^+$	$\bar{A}$	$\widetilde{A}$	$\bar{A}_P^R$	$\widetilde{A}_P$	$\bar{\bar{A}}_1$	$\bar{\bar{A}}_2$
4	0.01	0.03	3.46	0.81	0.53	3.21	3.21	0.76	0.49
		0.05	3.52	0.86	0.54	3.24	3.24	0.81	0.50
	0.02	0.03	3.62	0.87	0.56	3.34	3.34	0.81	0.50
		0.05	3.58	0.85	0.55	3.30	3.29	0.79	0.50
8	0.01	0.03	5.14	1.57	0.76	4.62	4.62	1.41	0.68
		0.05	5.04	1.62	0.77	4.52	4.48	1.46	0.69
	0.02	0.03	5.49	1.79	0.81	4.98	4.98	1.64	0.73
		0.05	5.35	1.73	0.80	4.77	4.75	1.53	0.73

we define $\bar{\bar{A}}_2$ as the average of the 2^{m+1} realizations obtained from $\boldsymbol{X}$ using the FQMC estimator together with the reversed permutation estimator. For example, for $m = 2$,

$$\bar{\bar{A}}_1 = \frac{A_2(X_1, X_2) + A_2(-X_1, -X_2) + A_2(X_2, X_1) + A_2(-X_2, -X_1)}{4}$$

and

$$\begin{aligned}\bar{\bar{A}}_2 = \frac{1}{8}\Big[&\Big(A_2(X_1, X_2) + A_2(X_1, -X_2) + A_2(-X_1, X_2) + A_2(-X_1, -X_2)\Big)\\ &+\Big(A_2(X_2, X_1) + A_2(X_2, -X_1) + A_2(-X_2, X_1) + A_2(-X_2, -X_1)\Big)\Big].\end{aligned}$$

Example 7 We make one final visit to our running bellwether example, where we note that the estimators $\bar{\bar{A}}_1$ and $\bar{\bar{A}}_2$ already appear in Table 7, and that the estimators $\bar{\bar{C}}_1$ and $\bar{\bar{C}}_2$ already reside in Table 8.

For the small values of m that we examined, the combined estimators $\bar{\bar{A}}_1$ and $\bar{\bar{A}}_2$ in Table 7 produced modest variance reductions of about 10% compared to the antithetic estimator $\bar{A}$ and the FQMC estimator $\widetilde{A}$, respectively (i.e., the corresponding estimators before the incorporation of the reversed permutation). The analogous improvements for the option-based estimators in Table 8 were slightly more marked. $\square$

Table 8: Standard errors ($\times 10^3$) of the naive, antithetic, FQMC, permutation, and combination estimators of the undiscounted option price $\mathrm{E}[(A_m^+ - k)^+]$ based on $n =$ 10,000 independent replications. For all cases, $s_0 = k = T = 1$.

m	ϕ	τ^2	s.e.						
			$\bar{C}^+$	$\bar{C}$	$\widetilde{C}$	$\bar{C}_P^R$	$\widetilde{C}_P$	$\bar{\bar{C}}_1$	$\bar{\bar{C}}_2$
4	0.01	0.03	2.68	1.42	0.89	2.46	2.46	1.26	0.77
		0.05	2.72	1.48	0.90	2.51	2.50	1.32	0.78
	0.02	0.03	2.89	1.49	0.93	2.65	2.65	1.32	0.79
		0.05	2.86	1.47	0.92	2.61	2.60	1.29	0.78
8	0.01	0.03	4.33	2.28	1.06	3.85	3.83	1.96	0.90
		0.05	4.21	2.31	1.08	3.75	3.69	1.99	0.91
	0.02	0.03	4.74	2.49	1.11	4.30	4.28	2.19	0.96
		0.05	4.60	2.43	1.11	4.08	4.03	2.08	0.95

2.4.4 Recapitulation

By extending the antithetic variates techniques to multivariate situations and taking advantage of the symmetry which results from this, we have proposed a full quasi-Monte Carlo technique which provides a notable reduction in variance while, under the right conditions, only moderately increasing computational time. For situations in which this method is not applicable, we have further provided a compromise between the naive estimator and our full quasi-Monte Carlo method, namely, our partial quasi-Monte Carlo estimator, which achieves much of the reduction in variance without increasing computational time to unreasonable levels. In addition, we list several related topics of interest which will be addressed later in the thesis.

1. Extensions to nonnormal increments. What happens if the increments $X_1, X_2, \ldots, X_m$ are i.i.d., but not normal? If they are symmetric around 0, then the easy-to-compute Equation (23) still holds. If the X_i's are not symmetric around 0, then we can use an inverse cumulative distribution function transformation as a work-around and retain the ability to use Equation (23). In any case, we conduct a small robustness study in §3.5 of the thesis to ascertain

the consequences of violations of the normality assumption.

2. We present more numerical results in §3.4 to further illustrate the effectiveness of our methodology as well as investigate the performance of our Gram–Charlier based p.d.f.'s as the number of replications increases.

3. Harmonic average estimators have also found their way into the literature. The harmonic average is simply the reciprocal of the arithmetic average of the reciprocals and is actually not too difficult to deal with, given what we have done so far. As a trivial example, suppose that X_1 and X_2 are i.i.d. Nor(0,1) increments, and $Y_i \equiv e^{X_i}$, $i = 1, 2$, are the associated lognormals. Then the Asian arithmetic average is $A = \frac{1}{2}(Y_1 + Y_1Y_2)$, and the harmonic average is

$$H = \frac{2}{\frac{1}{Y_1} + \frac{1}{Y_1Y_2}} \sim \frac{2}{Y_1 + Y_1Y_2} = 1/A,$$

since $e^{X_i} \sim e^{-X_i}$.

CHAPTER III

GRAM–CHARLIER PRICING OF ASIAN AVERAGES AND OPTIONS

This chapter details the construction of the Gram–Charlier (GC) estimator of a probability density function (p.d.f.) for the Asian arithmetic average based on a driving geometric Brownian motion (GBM) process such as that described in Chapter 2. The idea is to approximate the true p.d.f. of the Asian arithmetic average by a GC distribution, and then to use properties of the GC distribution as surrogates for the analogous true Asian arithmetic average properties. In particular, we will approximate the moments of the Asian average and the moments of associated options via the GC distribution, which we can obtain in more-or-less closed form. We follow closely the methodology laid out in Popovic and Goldsman [14].

The chapter is organized as follows. §3.1 gives relevant background material on GC methods. §3.2 concerns our implementation strategy, §3.3 discusses some efficiency tricks with respect to implementation, and §3.4 presents various motivational examples.

3.1 Background

Suppose that X has the p.d.f. denoted by $f_X(\cdot)$, and that this p.d.f. can be represented in terms of what is known as a Gram–Charler Type A series (Berberan-Santos [4]),

$$f_X(\xi) \;\equiv\; \phi(\xi)\sum_{i=0}^{\infty}\frac{\int_{-\infty}^{\infty}\varphi_i(x)f_X(x)\,\mathrm{d}x}{i!}\,\varphi_i(\xi) \;=\; \phi(\xi)\sum_{i=0}^{\infty}\frac{\mathrm{E}[\varphi_i(X)]}{i!}\,\varphi_i(\xi), \quad \xi\in\mathbb{R}, \tag{33}$$

where $\phi(\xi) \equiv \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\xi^2/2}$ is the standard normal p.d.f. and $\varphi_i(x)$ denotes the ith Hermite polynomial (see, for example, Gradshteyn and Ryzhik [9]). In fact, since

$$\{\varphi_0(x), \varphi_1(x), \varphi_2(x), \varphi_3(x), \varphi_4(x), \dots\} = \{1, x, x^2 - 1, x^3 - 3x, x^4 - 6x^2 + 3, \dots\}, \tag{34}$$

the $\mathrm{E}[\varphi_i(X)]$, $i = 1, 2, \dots$, are polynomials in the raw moments of X and can just be regarded as numbers once the expectations are taken. This means that Equation (33) is a polynomial (except for the factor $\phi(\cdot)$).

Assuming that all of the necessary moments exist and that the infinite summation converges nicely, Equation (33) shows that the p.d.f. $f_X(x)$ can be written as a standard normal reference p.d.f. multiplied by the GC Type A "correction," which accounts for skewness, kurtosis, and other higher-order moments of X relative to a standard normal.

3.2 Implementation

Unfortunately, the problem is that $f_X(\cdot)$ is unknown (for instance, in the case of an arithmetic Asian average, which has no closed form). The good news is that we can construct from Equation (33) an estimate of $f_X(\cdot)$, denoted by $\widehat{f}_X(\cdot)$, say. Following Popovic and Goldsman [14], suppose that we can simulate the random variable X, and that we conduct n replications of the simulation to obtain realizations $\boldsymbol{x} \equiv (x_1, x_2, \dots, x_n)$. We then calculate the corresponding sample moments $\frac{1}{n}\sum_{j=1}^{n} x_j^{\ell}$, $\ell = 0, 1, 2, \dots$, and use these to estimate $\mathrm{E}[\varphi_i(X)]$ (which is the expectation of a polynomial in X). Specifically, we plug the realizations (or, equivalently, the calculated sample moments) into Equations (33) and (34) to obtain the estimated p.d.f.

$$\widehat{f}_X(\xi) \equiv \phi(\xi) \sum_{i=0}^{\infty} \frac{\frac{1}{n}\sum_{j=1}^{n} \varphi_i(x_j)}{i!} \varphi_i(\xi) = \phi(\xi) \sum_{i=0}^{\infty} \frac{\overline{\varphi}_i(\boldsymbol{x})}{i!} \varphi_i(\xi), \quad \xi \in \mathbb{R},$$

where $\overline{\varphi}_i(\boldsymbol{x}) \equiv \frac{1}{n}\sum_{j=1}^{n} \varphi_i(x_j)$, $i = 0, 1, \dots$, is an infinite sequence of Hermite polynomials, but now expressed in terms of the raw sample moments.

There is still the issue of calculating an infinite number of moments. In addition to possible convergence problems of the infinite series in Equation (33), it is a fact that calculated raw sample moments sometimes exhibit standard errors that grow rapidly as the index ℓ of the moment increases — particularly in cases involving Asians having a large volatility parameter σ. This problem can be mitigated to some extent through the use of effective variance reduction techniques that we will consider in this chapter. Our examples use the first $\ell = 6$ empirically generated moments. The corresponding estimated p.d.f. generally seems to do a very good job of approximating the true distributional properties of the Asian arithmetic average.

The GC modeling methodology for averages turns out to work better if the underlying sample data is at least approximately similar to the normal reference density used in the GC Type A series. Can we nudge the data in the proper direction? In practice, we can help the p.d.f. estimate $\widehat{f}_X(\cdot)$ of X along by applying appropriate transformations to X so that the corresponding histogram resulting from a sample of realizations $\boldsymbol{X}$ roughly resembles a normal p.d.f. To this end, for example, we could take $Y = \ell\text{n}(X)$, so that Y is the log-transformation (or some other useful transformation) random variable with p.d.f. $f_Y(\cdot)$. Using Y, we calculate the raw sample moments of the data $\boldsymbol{y} \equiv (y_1, y_2, \ldots, y_n)$, where y_j is the outcome of the jth realization of Y. On plugging these sample moments into the estimator, we obtain the GC estimate of the p.d.f. $f_Y(\cdot)$,

$$\widehat{f}_Y(\xi) \;\equiv\; \phi(\xi) \sum_{i=0}^{q} \frac{\overline{\varphi}_i(\boldsymbol{y})}{i!}\, \varphi_i(\xi), \quad \xi \in \mathbb{R}, \tag{35}$$

where, typically, $q = 4$, 5, or 6. Then use the inverse transformation, relative to that detailed above, to obtain the target p.d.f. estimate, e.g., $\widehat{f}_X(y) = \widehat{f}_Y(\ell\text{n}(y))\frac{1}{y}$.

We summarize below the procedure for constructing the approximate GC distributions of the Asian arithmetic averages.

GC Probability Density Function Algorithm

1. Simulate n realizations of the GBM from Equation (1).

2. Use the realizations to construct n realizations of the Asian arithmetic average, $X_j = A_{m,j}$, $j = 1, 2, \ldots, n$.

3. If preliminary histograms do a poor job of conforming to the normal reference density, make an appropriate data transformation, $g(\cdot)$, say. Then for each realization $j = 1, 2, \ldots, n$, obtain $Y_j \equiv g(A_{m,j})$.

4. Calculate the sample moments of the random variable from the transformation Y and use them as inputs to the GC estimator $\widehat{f}_Y(\cdot)$.

5. Finally, obtain the GC estimate of the Asian's p.d.f. $f_X(x)$ by untransforming: $\widehat{f}_X(x) = \widehat{f}_Y(g(x))|g'(x)|$.

We close this subsection with two pertinent remarks on how our research plan plays into the algorithm. First of all, since the Gram–Charlier technique performs best when the target distribution in question is approximately standard normal, we use an obvious transformation on our observed data in Step 3 of the algorithm. Namely, let $\boldsymbol{A} \equiv (A_{m,1}, A_{m,2}, \ldots, A_{m,n})$ and define the function

$$Y \;=\; g(X|\boldsymbol{A}) \;=\; \frac{\ell\text{n}(X) - \frac{1}{n}\sum_{j=1}^{n} \ell\text{n}(A_j)}{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left[\ell\text{n}(A_j) - \frac{1}{n}\sum_{i=1}^{n}\ell\text{n}(A_i)\right]^2}}, \tag{36}$$

which gives the standardized versions of the natural logarithms of the original samples.

Second, there is nothing stopping us from using non-independent realizations X_j in Step 2 of the algorithm, so long as the realizations have the same distribution. Notably, even if the realizations $A_{m,1}, A_{m,2}, \ldots, A_{m,n}$ are correlated (hopefully negatively), the standardization obtained via Equation (36) is still roughly normal, if not quite of variance 1 — though it can be shown that the expected value of the sample variance of correlated but identically distributed data does indeed eventually converge to the underlying variance. In any case, this is precisely what we worked so hard to do in Chapter 2 of this thesis — generate such realizations efficiently and then use them in variance reduction schemes.

3.3 Implementation in Matrix Form

The purpose of this section is to give some details on efficient implementation of the GC methodology. Note that the results of this section produce the same numerical results as the methodology outlined in §3.2, but could potentially result in easier implementation of the GC method. With this goal in mind, define $\boldsymbol{B}$ as the matrix of constants corresponding to the coefficients of the first few Hermite polynomials given in Equation (34),

$$\boldsymbol{B} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -3 & 0 & 1 & 0 & 0 & 0 \\ 3 & 0 & -6 & 0 & 1 & 0 & 0 \\ 0 & 15 & 0 & -10 & 0 & 1 & 0 \\ -15 & 0 & 45 & 0 & -15 & 0 & 1 \end{bmatrix},$$

so that B_{ij} is the coefficient of the jth order term of the ith Hermite polynomial, for $i, j = 0, 1, \ldots, 6$.

Recall from §3.2 that we may utilize a data transformation $y = g(x)$, for example (see Equation (36)),

$$y_i \;=\; g(x_i) \;\equiv\; \frac{w_i - \bar{w}}{s_w}, \quad i = 1, 2, \ldots, n,$$

where $w_i \equiv \ell\text{n}(x_i)$, $i = 1, 2, \ldots, n$, $\bar{w} \equiv \sum_{i=1}^{n} w_i/n$ is the sample mean of the w_i's, and $s_w^2 \equiv \sum_{i=1}^{n}(w_i - \bar{w})^2/(n-1)$ is the sample variance.

With the transformed data in hand, now let $\boldsymbol{a}$ denote the vector of coefficients of the $\varphi_i(\xi)$ terms found in Equation (35),

$$\boldsymbol{a} \equiv \left[\frac{\overline{\varphi}_0(\mathbf{y})}{0!}, \frac{\overline{\varphi}_1(\mathbf{y})}{1!}, \ldots, \frac{\overline{\varphi}_6(\mathbf{y})}{6!}\right]^T,$$

where $\overline{\varphi}_i(\mathbf{y}) \equiv \frac{1}{n}\sum_{j=1}^{n} \varphi_i(y_j)$, $i = 0, 1, \ldots, 6$; thus, these coefficients are themselves Hermite polynomials derived from the sample moments of $y_1, y_2, \ldots, y_n$.

In order to eventually obtain the GC estimate of the Asian's p.d.f. $f_X(x)$ by untransforming from $g(x) = (\ell\text{n}(x) - \bar{w})/s_w$, we have

$$\begin{aligned}
\widehat{f}_X(x) &= \widehat{f}_Y(g(x))|g'(x)| \\
&= \frac{1}{s_w x}\phi(g(x))\boldsymbol{B}\boldsymbol{a}\left[1,\ g(x),\ \ldots,\ (g(x))^6\right] \\
&= \frac{1}{s_w x}\phi(g(x))\sum_{i=0}^{6} c_i (g(x))^i,
\end{aligned}$$

where $\boldsymbol{c} = [c_0, c_1, \ldots, c_6]^T \equiv \boldsymbol{B}\boldsymbol{a}$.

3.4 Examples

To demonstrate how the Gram–Charlier method works, and then to see how we can implement variance reduction schemes, we performed a series of Monte Carlo experiments.

To begin with, we conducted a baseline MC experiment involving 10,000,000 independent replications for the purpose of establishing a "perfect" benchmark p.d.f. that we can use to compare against estimated p.d.f.'s obtained via GC. In our example, we set $m = 32$, $s_0 = 1$, $r = 0.05$, and $\sigma = 0.1$; henceforth, we will call this our "standard" parameter configuration for our subsequent examples. Figure 3 depicts our "perfect" GC fit overlaid with the histogram of the actual replications. As is easily seen, this fit indeed deserves the moniker "perfect" p.d.f. — which makes sense since it is based on so many replications. We will use this bellwether distribution as a basis of comparison among the various estimated p.d.f.'s that we will examine in the sequel.

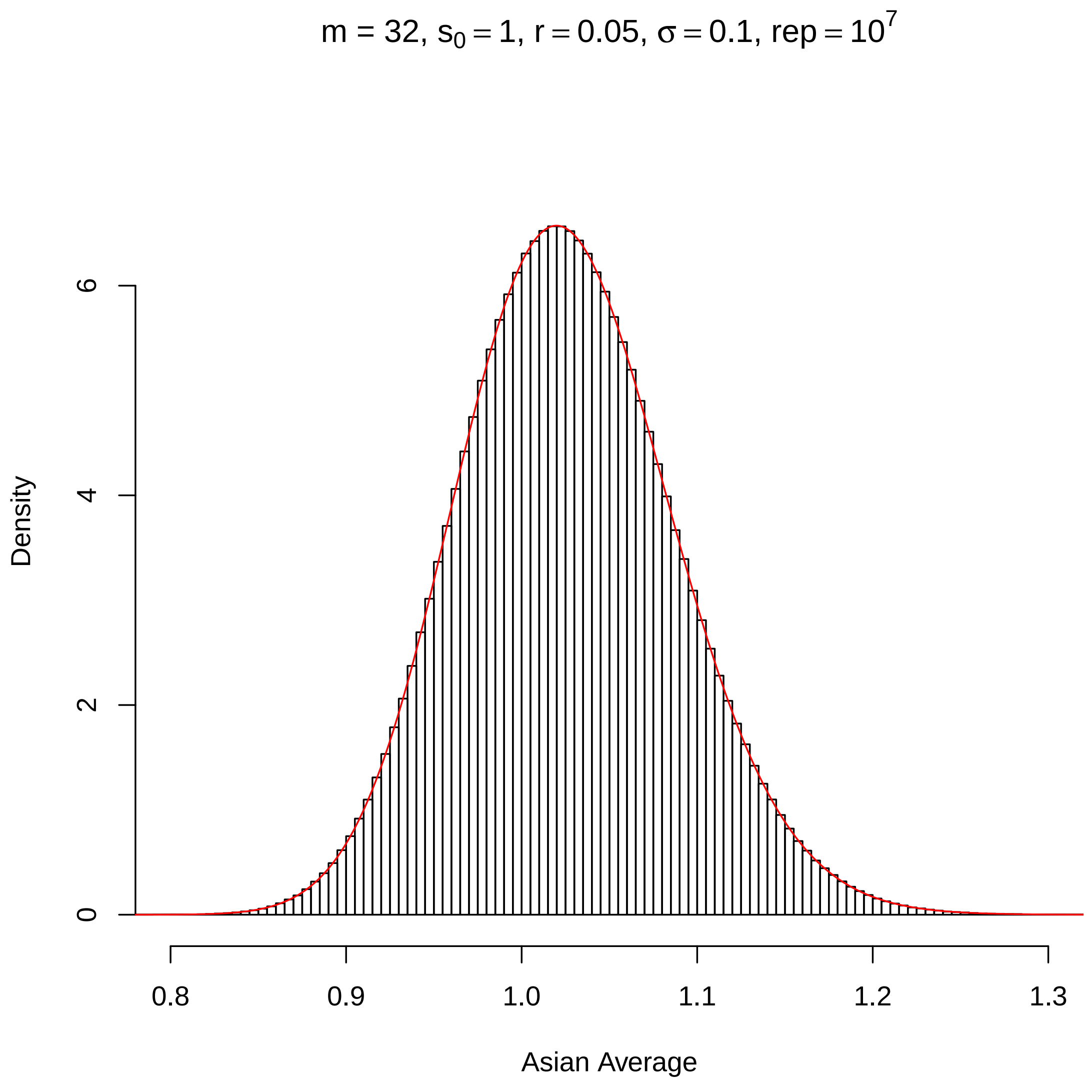

Figure 3: Gram–Charlier "perfect" fit for the standard parameter configuration

Since we often need to perform a potentially computationally intensive transformation on the Asian average in order to estimate the p.d.f. using Gram–Charlier, and since we have anecdotal evidence that the full FQMC does not yield variance reductions that are substantially better than those of the partial PQMC method, our subsequent examples in this section incorporate PQMC.

Example 8 With an eye on eventually using the GC method, we are interested in estimating the first 6 moments of the Asian arithmetic average. We use method of moments (MoM) estimators, which are unbiased for the rth moment, $\mathrm{E}[(A_m^+)^r]$. For instance, we can estimate the rth moment by the naive MoM estimator, $\frac{1}{n}\sum_{j=1}^n A_{m,j}^r$. Similar remarks hold for the corresponding antithetic and PQMC estimator incarnations. If our GC algorithm happens to require some transformation $Y = g(A_m)$, we base our MoM estimators on the resulting Y_j's instead of on the original $A_{m,j}$'s.

Now the task at hand simply amounts to comparing unbiased estimators on the basis of their variances — smaller is better. Specifically, we compare the variances of the naive, antithetic, and PQMC moment estimators, the latter of which uses minimum block size of 4 (in the parlance of the previous chapter). To do so, we performed 100,000 independent replications using $s_0 = 1$, $m = 32$, and various parameter settings. Table 9 summarizes the results. The bottom line is that the PQMC method consistently delivers unbiased moment estimators having the smallest variance; so it is these estimators that we will use as plug-ins for our GC p.d.f. estimation algorithm. $\square$

Table 9: Sample variances of Asian arithmetic average moment estimators using the naive, antithetic, and PQMC (minimum block size of 4) estimators based on $n = 100{,}000$ independent replications. For all cases, $s_0 = 1$ and $m = 32$.

r	σ^2	naive estimator moments						antithetic estimator moments						PQMC estimator moments					
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
0.04	0.04	0.015	0.07	0.17	0.36	0.7	1.4	0.0002	0.002	0.01	0.04	0.1	0.3	3E-5	0.0005	0.003	0.01	0.03	0.1
	0.08	0.030	0.14	0.40	1.01	2.5	6.3	0.0006	0.009	0.05	0.20	0.7	2.1	14E-5	0.0020	0.011	0.04	0.14	0.4
	0.12	0.045	0.23	0.73	2.18	6.8	23.8	0.0014	0.021	0.13	0.58	2.9	9.0	31E-5	0.0047	0.028	0.12	0.48	1.8
	0.16	0.061	0.32	1.16	4.12	16.8	86.1	0.0025	0.040	0.27	1.42	7.8	39.9	56E-5	0.0088	0.057	0.29	1.37	7.3
0.06	0.04	0.015	0.07	0.18	0.40	0.8	1.6	0.0002	0.002	0.01	0.04	0.1	0.3	4E-5	0.0005	0.003	0.01	0.03	0.1
	0.08	0.031	0.15	0.43	1.11	2.7	6.9	0.0006	0.010	0.06	0.22	0.8	2.4	14E-5	0.0021	0.012	0.05	0.16	0.5
	0.12	0.046	0.23	0.76	2.26	6.9	22.2	0.0014	0.022	0.14	0.61	2.4	9.3	32E-5	0.0049	0.030	0.13	0.53	2.0
	0.16	0.062	0.33	1.23	4.41	17.3	78.0	0.0025	0.041	0.28	1.43	6.9	34.6	57E-5	0.0090	0.060	0.30	1.46	7.6
0.08	0.04	0.016	0.07	0.19	0.44	0.9	1.8	0.0002	0.002	0.01	0.05	0.2	0.4	4E-5	0.0005	0.003	0.01	0.03	0.1
	0.08	0.031	0.15	0.46	1.22	3.1	8.4	0.0006	0.010	0.06	0.24	0.9	2.8	15E-5	0.0022	0.013	0.05	0.18	0.6
	0.12	0.048	0.25	0.85	2.65	8.6	31.0	0.0015	0.024	0.15	0.72	3.0	12.2	33E-5	0.0052	0.033	0.15	0.60	2.4
	0.16	0.064	0.35	1.31	4.63	17.5	73.6	0.0025	0.042	0.29	1.53	7.5	38.9	58E-5	0.0094	0.064	0.33	1.59	8.3
0.10	0.04	0.016	0.08	0.21	0.48	1.0	2.0	0.0002	0.003	0.01	0.05	0.2	0.4	4E-5	0.0006	0.003	0.01	0.03	0.1
	0.08	0.033	0.16	0.50	1.32	3.4	8.9	0.0007	0.011	0.07	0.27	1.0	3.3	15E-5	0.0023	0.014	0.06	0.20	0.7
	0.12	0.049	0.26	0.90	2.88	9.6	35.7	0.0015	0.025	0.16	0.78	3.3	14.1	34E-5	0.0054	0.035	0.16	0.66	2.7
	0.16	0.067	0.38	1.45	5.48	23.3	117.7	0.0026	0.046	0.33	1.79	9.4	52.1	60E-5	0.0100	0.069	0.37	1.83	9.7
0.12	0.04	0.016	0.08	0.22	0.52	1.1	2.3	0.0002	0.003	0.02	0.06	0.2	0.5	4E-5	0.0006	0.003	0.01	0.04	0.1
	0.08	0.033	0.17	0.52	1.42	3.7	10.1	0.0007	0.011	0.07	0.28	1.0	3.3	15E-5	0.0024	0.015	0.06	0.22	0.7
	0.12	0.050	0.27	0.95	3.07	10.2	36.9	0.0015	0.026	0.17	0.82	3.5	14.4	34E-5	0.0056	0.037	0.17	0.72	2.9
	0.16	0.068	0.39	1.54	5.91	25.4	132.0	0.0027	0.047	0.34	1.89	9.9	55.3	62E-5	0.0107	0.076	0.42	2.40	17.9

Example 9 To study the performance of our GC p.d.f. estimation algorithm, we performed five sets of runs on the standard configuration problem, i.e., $m = 32$, $s_0 = 1$, $r = 0.05$, and $\sigma = 0.1$. The goal was to determine which estimators produce p.d.f.'s that best approximate the true p.d.f., in this case the "perfect" proxy p.d.f. of Figure 3; in addition, how many independent sample path replications were necessary to generate good p.d.f. approximations?

Figures 4–7 illustrate GC p.d.f.'s based on the naive and PQMC (minimum block size 4) moment estimators; the four figures correspond to sample runs of $n = 50$, 100, 250, and 500 replications, respectively. Each individual figure depicts five sample p.d.f.'s calculated from five different sets of n replications. In each case, the "perfect" p.d.f. is also overlaid in red as a basis for comparison. It is readily apparent that PQMC p.d.f. performance improves as the number of replications n increases; and the estimated PQMC p.d.f.'s seem to be stabilizing as we approach 500 replications. The convergence is not so clear for the naive p.d.f.'s. In any case, the PQMC plots exhibit less variability than the naive plots; and the PQMC p.d.f.'s are always closer to the "perfect" p.d.f. $\square$

Example 10 Another question of interest with respect to the Gram–Charlier method is that of choosing the number of estimated moments to utilize in the GC calculations — one would think that more is better, though more computationally costly. It turns out that as long as the experimenter succeeds in standardizing the Asian average into a close approximation of the standard normal distribution, the number of moments may not be such an overwhelming factor with regard to GC performance. We conducted an experiment with our standard parameter settings and $n = 500$ replications. In these runs, we utilized 2, 3, 4, 5, and 6 moments to derive our GC p.d.f. estimates. See the resulting p.d.f.'s in Figure 8, which are so close as to be hard to distinguish. Again, the PQMC p.d.f.'s dominate those of the naive estimator. $\square$

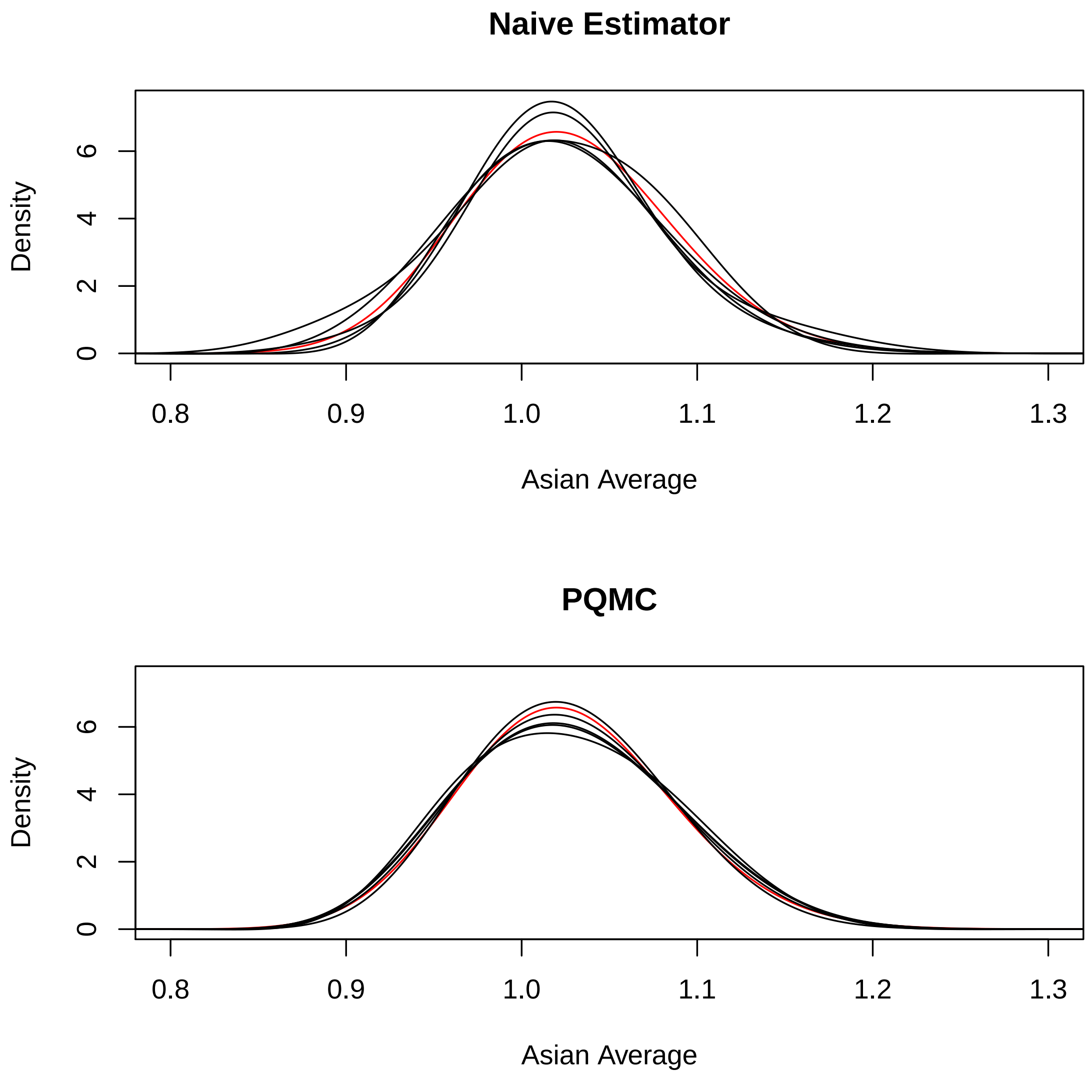

Figure 4: Five sets of simulations of the Gram–Charlier p.d.f., each using 50 replications and our standard parameter configuration. In this example, the PQMC method uses a minimum block size of 4. The red plot corresponds to the "perfect" proxy p.d.f. of Figure 3.

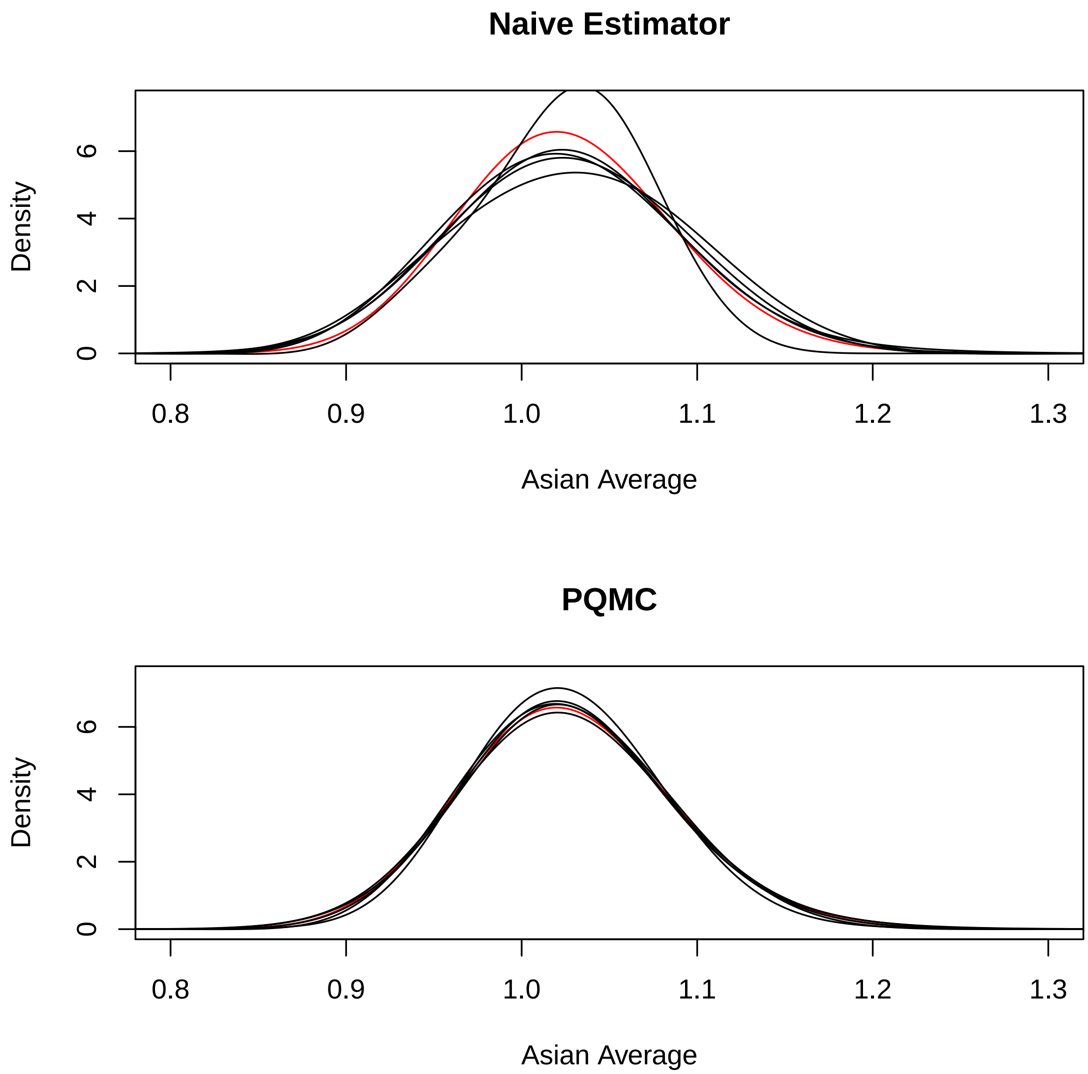

Figure 5: Five sets of simulations of the Gram–Charlier p.d.f., each using using 100 replications and our standard parameter configuration. In this example, the PQMC method uses a minimum block size of 4. The red plot corresponds to the "perfect" proxy p.d.f. of Figure 3.

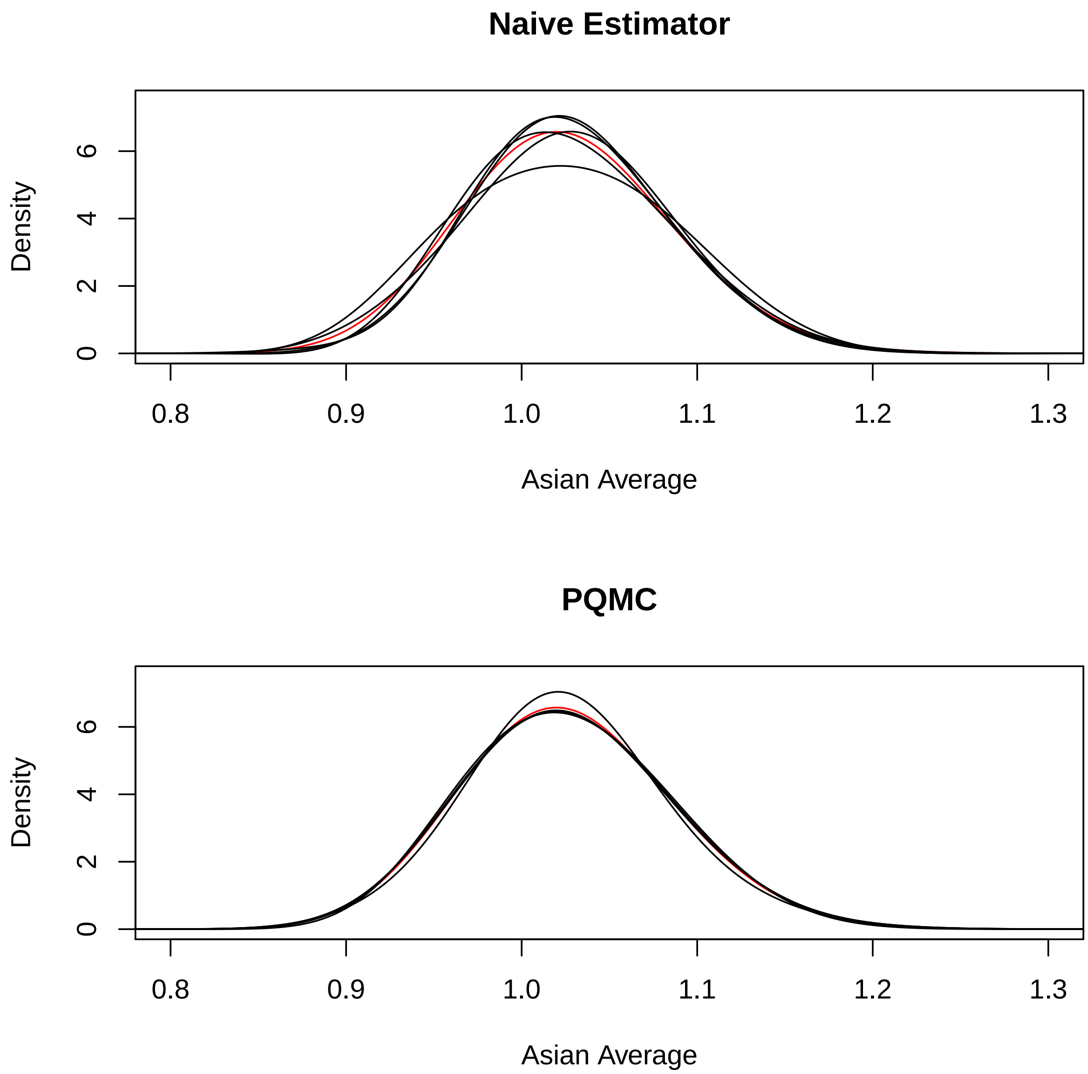

Figure 6: Five sets of simulations of the Gram–Charlier p.d.f., each using using 250 replications and our standard parameter configuration. In this example, the PQMC method uses a minimum block size of 4. The red plot corresponds to the "perfect" proxy p.d.f. of Figure 3.

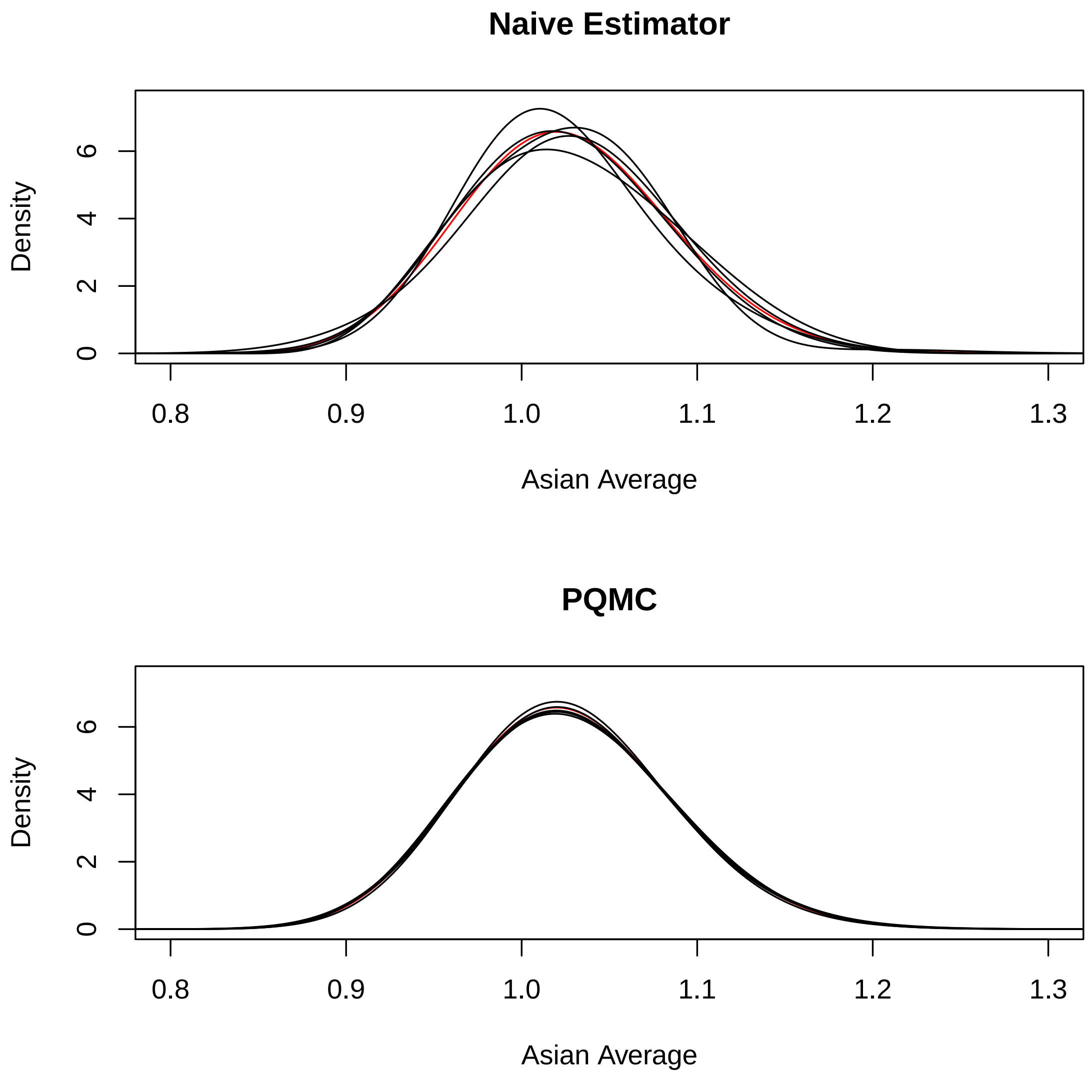

Figure 7: Five sets of simulations of the Gram–Charlier p.d.f., each using using 500 replications and our standard parameter configuration. In this example, the PQMC method uses a minimum block size of 4. The red plot corresponds to the "perfect" proxy p.d.f. of Figure 3.

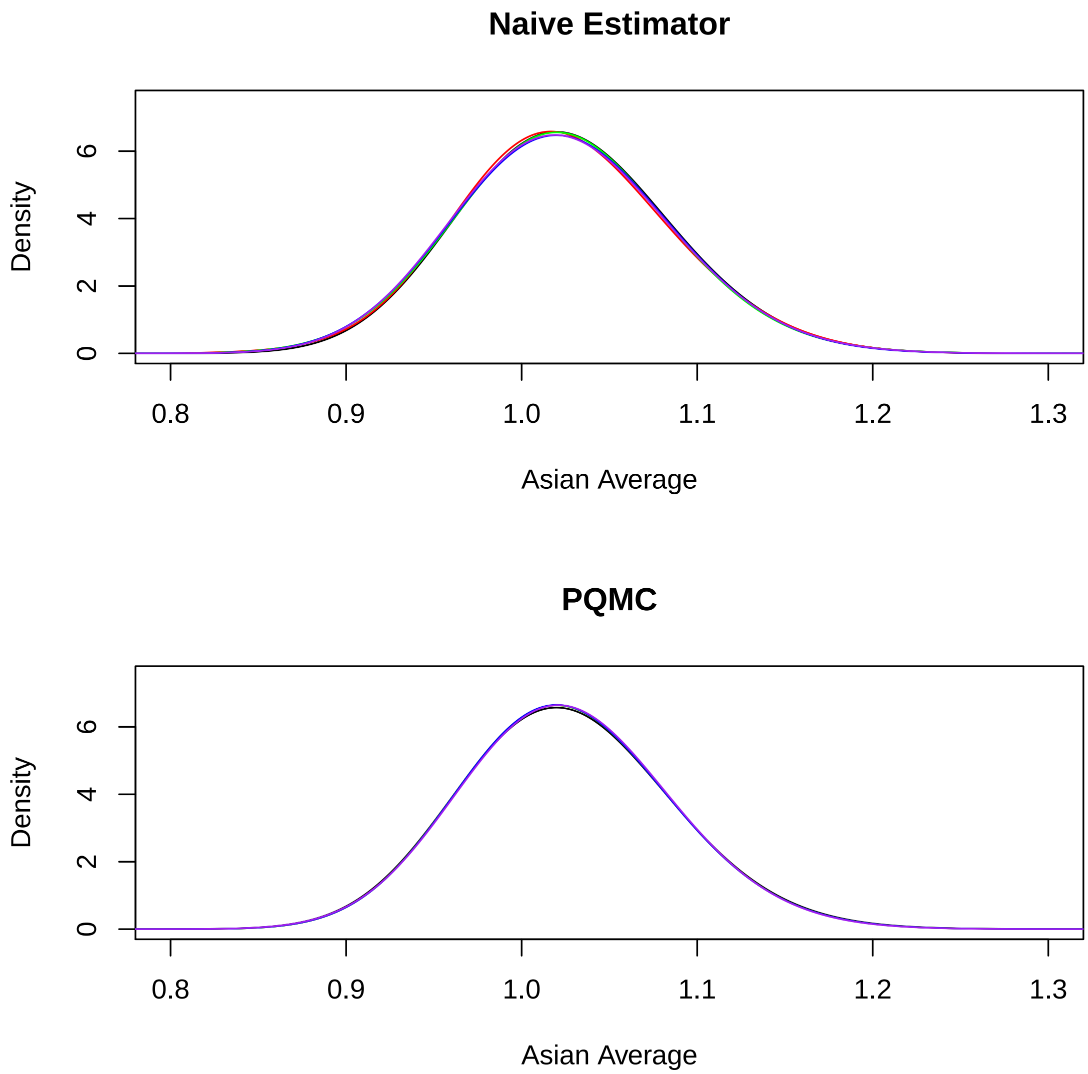

Figure 8: Simulations of the Gram–Charlier p.d.f. using 500 replications and 2, 3, 4, 5, and 6 moments. In this example, the PQMC method uses a minimum block size of 4. The red plot (though hard to see) corresponds to the "perfect" proxy p.d.f. of Figure 3.

Example 11 We performed a battery of simulations to study convergence and sensitivity properties of the GC coefficients $\boldsymbol{c}$ from §3.3. Representative results are presented in Tables 10–12. In particular, Table 10 illustrates how the number of replications n affects the convergence of the coefficients — at least for the special case $m = 32$, $r = 0.07$, and $\sigma = 0.1$. We see from the table that the values of $\bar{w}$, s_w^2, and $\boldsymbol{c} = (c_0, c_1, \ldots, c_6)$ all converge very rapidly as the number of replications increases; and the coefficients c_4, c_5, and c_6 are all nearly 0 by the time we get to large values of n. The purpose of Table 11 is to illustrate what happens to the GC coefficients as we increase m; and we again see rapid convergence. Table 12 provides sensitivity results as we vary r and σ — larger values of σ unsurprisingly result in larger s_w^2 as well as some larger coefficient values. $\Box$

Table 10: Coefficients of the Gram–Charlier p.d.f. for $m = 32$, $r = 0.07$, $\sigma = 0.1$, and various numbers of replications. All estimates are derived using the PQMC estimator with minimum block size of 4.

n	$\bar{w}$	s_w^2	c_0	c_1	c_2	c_3	c_4	c_5	c_6
50	0.035	0.0607	1.00902	−0.01763	−0.02067	0.00673	0.00635	−0.00017	−0.00049
100	0.035	0.0611	1.01361	−0.01895	−0.03359	0.00787	0.00980	−0.00031	−0.00063
250	0.035	0.0591	1.00821	−0.01732	−0.01976	0.00666	0.00547	−0.00018	−0.00032
500	0.035	0.0590	1.01875	−0.01626	−0.04215	0.00544	0.00938	−0.00000	−0.00032
1000	0.035	0.0590	1.00908	−0.01406	−0.01395	0.00392	0.00025	0.00015	0.00028
10000	0.035	0.0592	1.00382	−0.01604	−0.00815	0.00528	0.00160	0.00001	−0.00003
100000	0.035	0.0595	1.00023	−0.01602	0.00025	0.00531	−0.00039	0.00001	0.00005
1000000	0.035	0.0596	1.00012	−0.01601	−0.00015	0.00533	−0.00002	0.00000	0.00001

Table 11: Coefficients of the Gram–Charlier p.d.f. for 10^5 replications, $r = 0.07$, $\sigma = 0.1$, and various values of m. All estimates are derived using the PQMC estimator with minimum block size of 4.

m	$\bar{w}$	s_w^2	c_0	c_1	c_2	c_3	c_4	c_5	c_6
4	0.042	0.0688	0.99862	−0.00983	0.00341	0.00342	−0.00090	−0.00003	0.00004
8	0.038	0.0636	1.00027	−0.01288	−0.00032	0.00425	−0.00005	0.00001	0.00001
16	0.036	0.0609	1.00000	−0.01465	0.00129	0.00473	−0.00086	0.00003	0.00009
32	0.035	0.0595	1.00079	−0.01584	−0.00127	0.00519	0.00006	0.00002	0.00002
64	0.034	0.0590	1.00013	−0.01653	−0.00037	0.00547	0.00012	0.00001	−0.00001
128	0.034	0.0585	1.00057	−0.01685	−0.00102	0.00559	0.00011	0.00001	0.00001
256	0.034	0.0584	0.99974	−0.01683	0.00122	0.00556	−0.00055	0.00001	0.00005
512	0.034	0.0583	1.00123	−0.01723	−0.00353	0.00579	0.00112	−0.00001	−0.00007
1024	0.034	0.0583	0.99969	−0.01714	0.00081	0.00573	−0.00024	−0.00000	0.00001
2048	0.034	0.0582	0.99764	−0.01736	0.00549	0.00590	−0.00130	−0.00002	0.00005

Table 12: Coefficients of the Gram–Charlier p.d.f. for 10^5 replications, $m = 32$, and various levels of r and σ. All estimates are derived using the PQMC estimator with minimum block size of 4.

r	σ	$\bar{w}$	s_w^2	c_0	c_1	c_2	c_3	c_4	c_5	c_6
	0.05	0.025	0.0297	0.999	−0.00801	0.00116	0.00267	−0.00026	−0.00000	0.00001
	0.10	0.024	0.0594	0.999	−0.01594	0.00210	0.00530	−0.00087	0.00000	0.00007
0.05	0.15	0.022	0.0888	1.000	−0.02359	0.00222	0.00768	−0.00108	0.00004	0.00009
	0.30	0.010	0.1776	0.999	−0.04778	0.00485	0.01589	−0.00197	0.00001	0.00015
	0.50	−0.018	0.2933	0.998	−0.07916	0.01047	0.02619	−0.00467	0.00004	0.00039
	0.05	0.052	0.0299	0.999	−0.00804	0.00233	0.00269	−0.00061	−0.00000	0.00003
	0.10	0.052	0.0598	1.000	−0.01595	−0.00078	0.00532	0.00013	−0.00000	−0.00000
0.10	0.15	0.048	0.0896	0.999	−0.02374	0.00375	0.00784	−0.00143	0.00001	0.00011
	0.30	0.036	0.1785	1.001	−0.04711	−0.00077	0.01522	−0.00089	0.00010	0.00014
	0.50	0.008	0.2946	0.999	−0.07912	0.00692	0.02637	−0.00362	0.00000	0.00033
	0.05	0.078	0.0301	0.999	−0.00796	0.00133	0.00265	−0.00037	−0.00000	0.00002
	0.10	0.077	0.0602	1.000	−0.01590	0.00087	0.00529	−0.00036	0.00000	0.00003
0.15	0.15	0.074	0.0901	0.999	−0.02415	0.00159	0.00822	−0.00016	−0.00003	−0.00001
	0.30	0.062	0.1798	1.000	−0.04754	0.00182	0.01593	−0.00110	−0.00002	0.00011
	0.50	0.034	0.2968	1.000	−0.07893	0.00645	0.02640	−0.00338	−0.00002	0.00031

3.5 Non-Normal Increments

A problem of great interest concerns the case in which the increments of the driving stock price process are non-normal — the assumption of normal increments is certainly not guaranteed to be realistic in practice. In other words, instead of the classic process described by Equation (1), we consider one of the form

$$S(t) \equiv s_0 \exp\left\{ \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma \mathcal{Z}(t) \right\}, \quad t \geq 0, \tag{37}$$

where $\mathcal{Z}(t)$ is now assumed to have stationary and independent (but not necessarily i.i.d. normal) increments,

$$X_j \equiv \mathcal{Z}\left(\frac{jT}{m} \right) - \mathcal{Z}\left(\frac{(j-1)T}{m} \right), \quad j = 1, 2, \ldots, m. \tag{38}$$

If the increments are symmetric about zero, we can still implement all of the techniques outlined in previous sections of the thesis with almost no additional work — though calculation of quantities such as exact moments will require re-derivation. In fact, all we need to do with respect to implementation is to use the X_j's from Equation (38) as the new increments and proceed as in the i.i.d. normal increments case.

On the other hand, if the distribution of the i.i.d. increments is non-symmetric, one must resort to a small fix in order to implement variance reduction techniques such as antithetics and PQMC. The starting point is the well-known inverse transform theorem, which states that if X is a continuous random variable with c.d.f. $F(x)$, then $F(X) \sim \text{Unif}(0,1)$; and thus, if $U \sim \text{Unif}(0,1)$, then $X = F^{-1}(U)$ has the distribution possessing c.d.f. $F(x)$ (see Law [12] for a more-thorough discussion on random variate generation methodology). Moreover, with antithetic variance reduction in mind, we note that the pair of random variables $F^{-1}(U)$ and $F^{-1}(1-U)$ are often negatively correlated — whether or not the increments are symmetric.

To illustrate the effects of non-normal increments, we performed several simulation experiments using our "standard" parameter settings, but with Laplace, uniform, and shifted exponential increments instead of the bellwether normal increments. The Laplace p.d.f. is of the form $f(x) = \frac{\lambda}{2}e^{-\lambda|x|}$ for $x \in \mathbb{R}$ and $\lambda > 0$, and is therefore symmetric about 0; the uniform distribution is obviously symmetric; but the shifted exponential is skewed, having p.d.f. of the form $f(x) = \lambda e^{-\lambda(x-k)}$ for $x > 0$, $\lambda > 0$, and shift k. In particular, we used Laplace and uniform increments having mean 0 and variance τ^2. To generate pairs of negatively correlated shifted exponential increments with mean 0 and variance τ^2, we simply take $X = F^{-1}(U) = \tau(-\ell\text{n}(U) - 1)$ and $X' = F^{-1}(1-U) = \tau(-\ell\text{n}(1-U) - 1)$.

Table 13 gives Gram–Charlier coefficient estimates of the p.d.f.'s of the Asian average for the standard case $s_0 = 1$, $r = 0.05$, $\sigma = 0.1$, $m = 32$, and various distributions of the underlying increments; all results are based on $n = 10^5$ replications and utilize the PQMC method with minimum block size of 4. We see that the coefficients of the shifted exponential case are significantly different than those arising from the other (symmetric) distributions.

Table 13: Estimated coefficients of the Gram–Charlier p.d.f.'s of the Asian average for $r = 0.05$, $\sigma = 0.1$, $m = 32$, and various distributions of the underlying increments; all results are based on $n = 10^5$ replications and utilize the PQMC method with minimum block size of 4.

Distribution	c_0	c_1	c_2	c_3	c_4	c_5	c_6
Normal	1.000	−0.01599	0.00034	0.00532	−0.00016	0.00000	0.00001
Laplace	1.019	−0.01447	−0.03502	0.00374	0.00472	0.00022	0.00015
Uniform	0.991	−0.01660	0.01884	0.00595	−0.00346	−0.00008	0.00004
Shifted Exponential	1.001	−0.07832	0.00302	0.02546	−0.00308	0.00013	0.00034

Figures 9–12 depict the histograms and resulting Gram–Charlier p.d.f.'s of the Asian averages for the standard case $s_0 = 1$, $r = 0.05$, $\sigma = 0.1$, $m = 32$, in the presence of normal, Laplace, uniform, and shifted exponential increments, respectively. All increments have mean 0 and variance τ^2, a function of σ, T, and m as defined in §2.3, and all results are based on $n = 10^6$ independent replications using PQMC variance reduction (again with minimum block size of 4) to smooth out the curves. Figure 13

consolidates the four GC curves for easy apples-to-apples comparison, where we again see that the shifted exponential increments produce p.d.f.'s that differ significantly from the others. The figures show that, even though the increments driving the process are not from the normal distribution, with a little bit of forethought, our methodology can be applied to derive a GC p.d.f. estimate that does an admirable job of fitting the distribution of the new process.

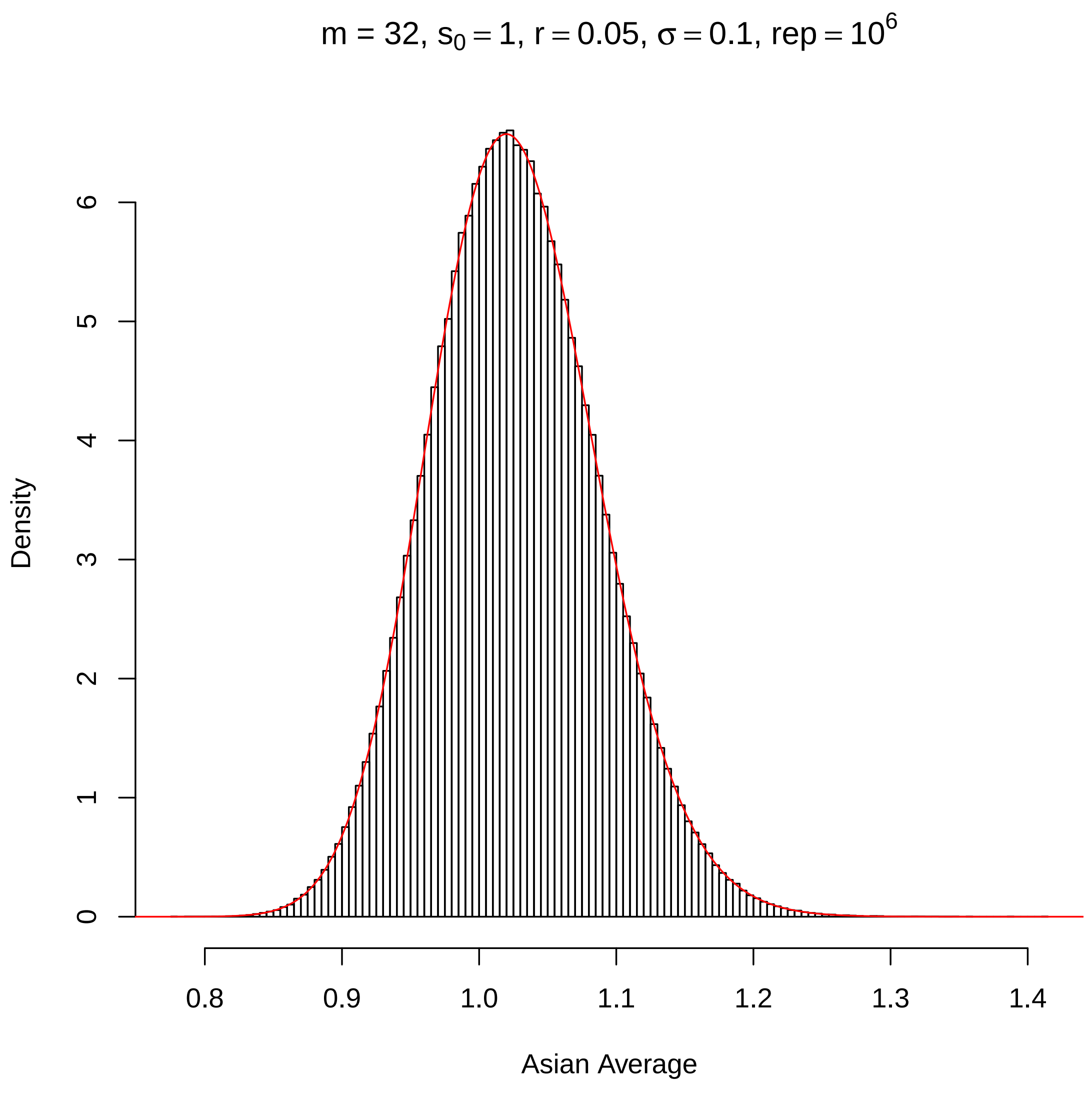

Figure 9: Simulations of the Gram–Charlier p.d.f. using 10^6 replications and normal increments

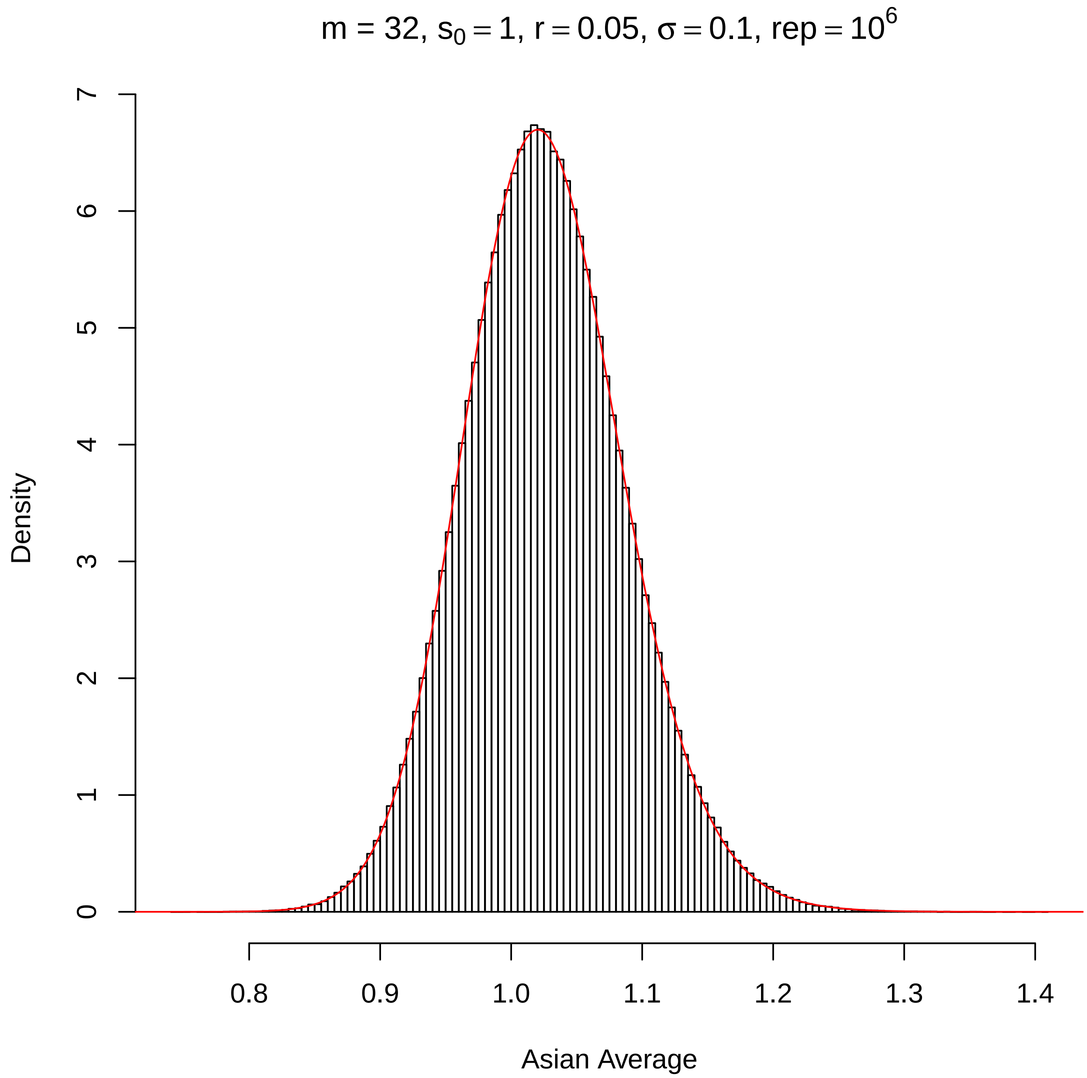

Figure 10: Simulations of the Gram–Charlier p.d.f. using 10^6 replications and Laplace increments

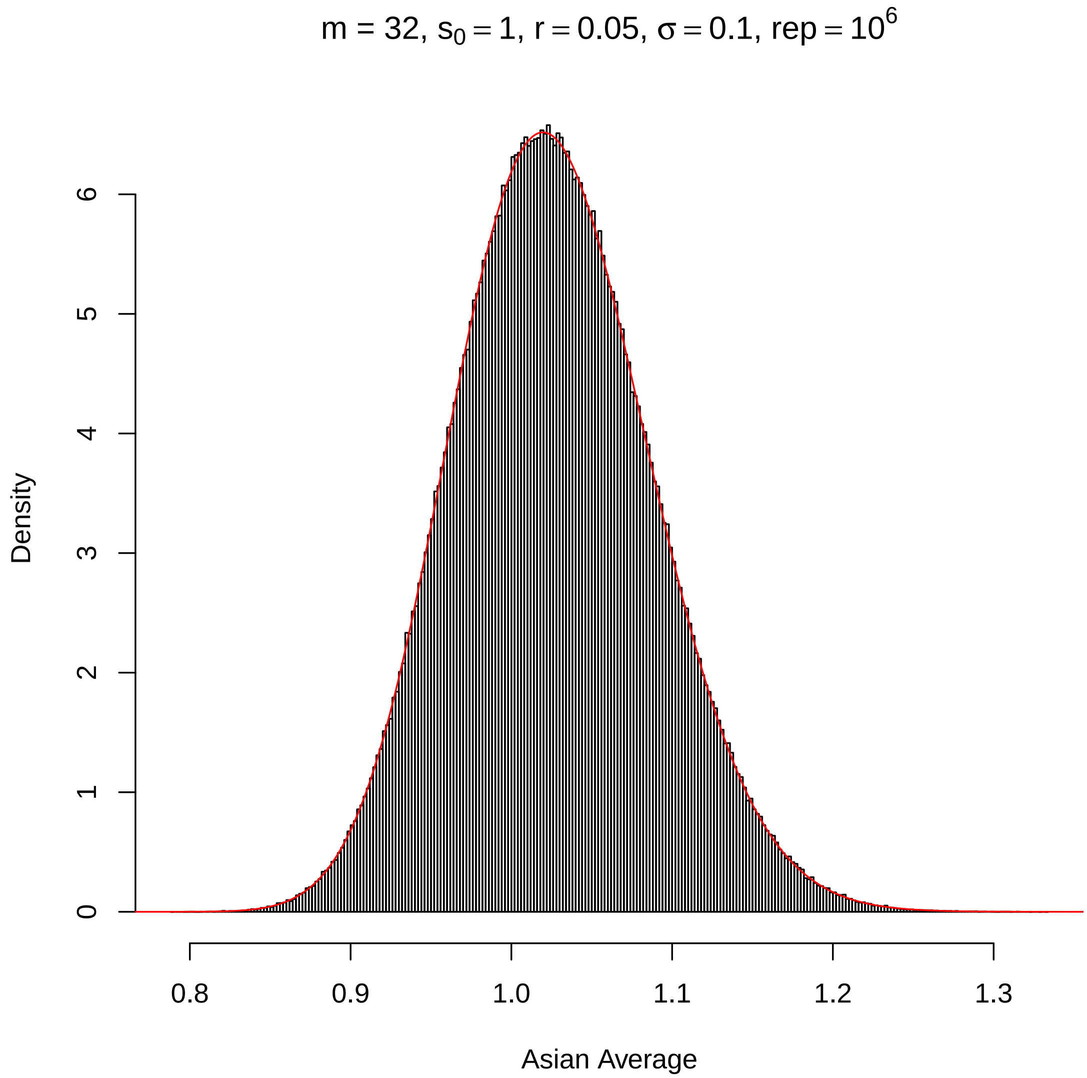

Figure 11: Simulations of the Gram–Charlier p.d.f. using 10^6 replications and uniform increments

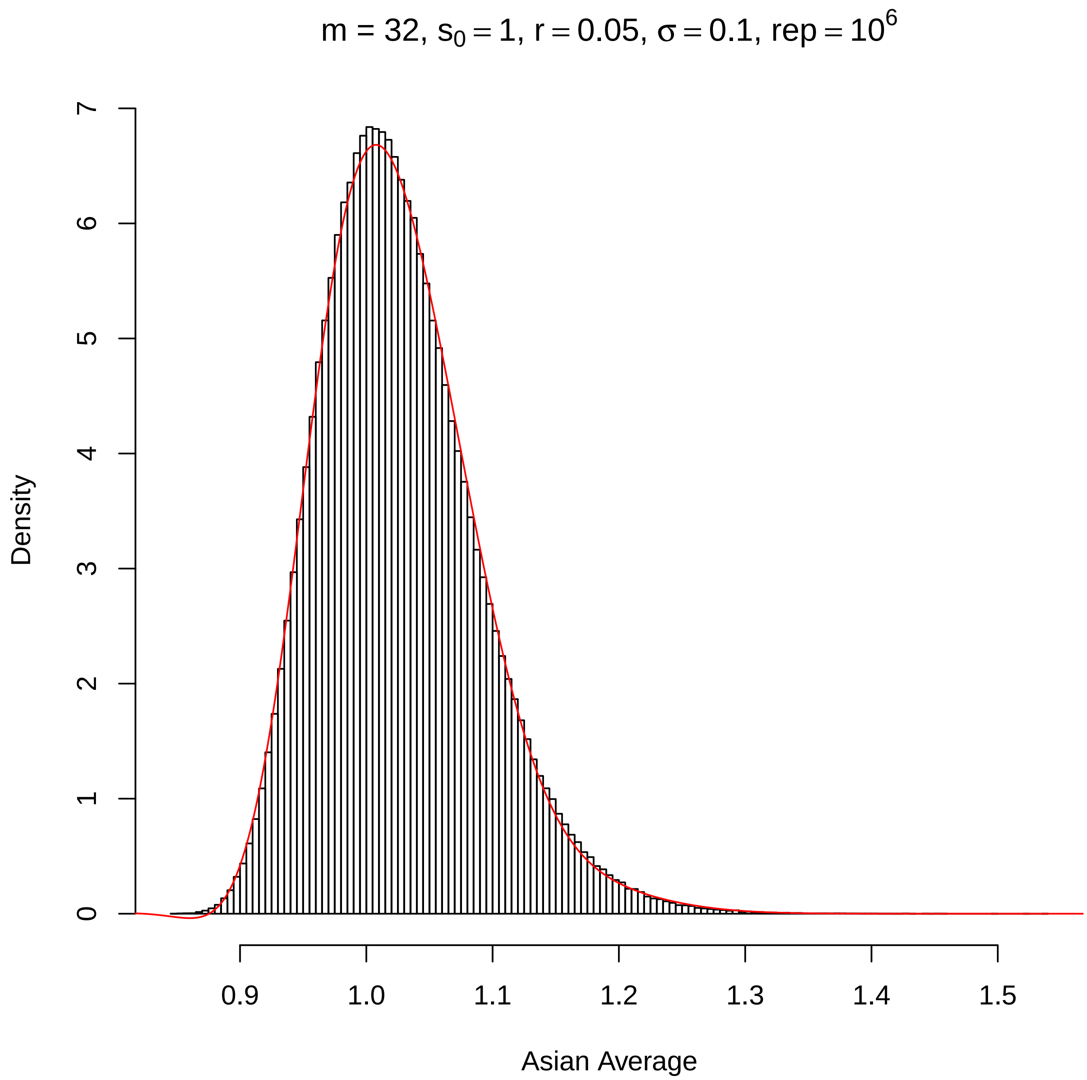

Figure 12: Simulations of the Gram–Charlier p.d.f. using 10^6 replications and shifted exponential increments

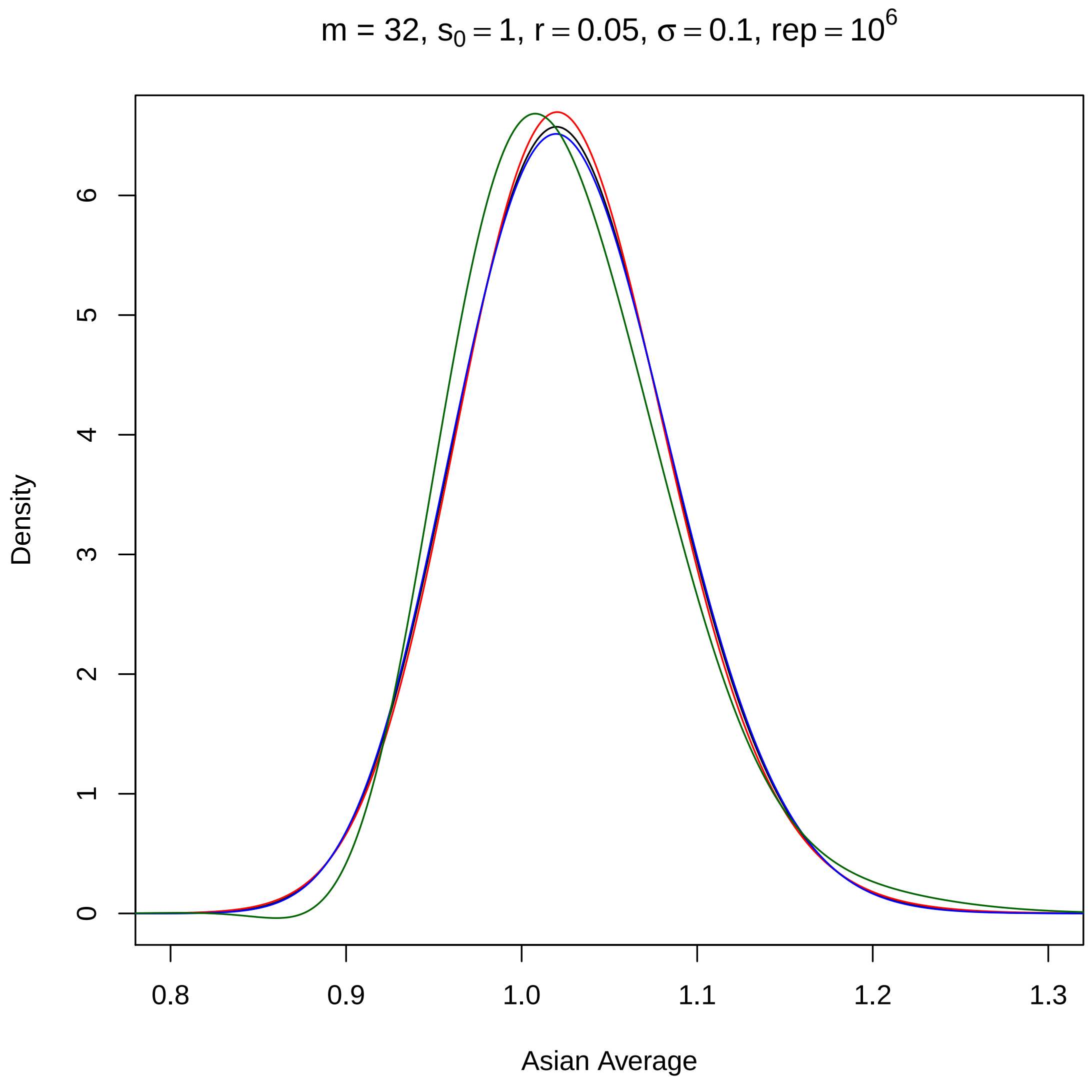

Figure 13: Comparison of the Gram–Charlier p.d.f.'s using 10^6 replications and various increments. The p.d.f. using normal increments is in black, Laplace is in red, uniform is in blue, and shifted exponential is in green.

CHAPTER IV

CONDITIONAL PROBABILITY OF CORRECT SELECTION FOR RANKING AND SELECTION PROCEDURES

This chapter is concerned with the post hoc analysis of various ranking and selection procedures. We investigate how procedure termination conditions of various single-stage and multi-stage procedures affect the probability of correct selection ($P\{\text{CS}\}$) — that is, how often a procedure correctly selects the best competing alternative. For example, consider a procedure designed to select the population having the largest mean while ensuring that the correct population is indeed chosen with a probability of at least 0.90. A scenario in which that procedure finishes after only a few stages is a much different than a scenario in which the procedure had continued for a large number of stages. The scenario which finished early might have a conditional $P\{\text{CS}\}$ (CPCS) much higher than 0.90 (i.e., it was extremely easy to select the best alternative population), while the scenario which continued for several more stages might well have a CPCS significantly less than 0.90 (because the task of finding the best population may have been more difficult in that case). By considering the stage at which the procedure terminated and the sample path which resulted in this termination, we can estimate the conditional probability of correct selection. This will allow the experimenter to make more-accurate statements than could be made by using the (unconditional) $P\{\text{CS}\}$ alone.

The organization of this chapter is as follows. §4.1 provides background material on ranking and selection's well-known indifference-zone approach. The remaining sections §§4.2–4.4 outline various R&S procedures and discuss how they perform in terms of conditional $P\{\text{CS}\}$.

4.1 Background

For now, suppose that we are interested in the archetypal ranking and selection problem — selecting that one of k normal populations having the largest mean. We are concerned with sequential procedures which use the indifference-zone approach to make such decisions.

By way of set-up, assume that independent and identically distributed (i.i.d.) observations $Y_{i1}, Y_{i2}, \ldots$ $(1 \leq i \leq k)$ are taken from $k \geq 2$ normal populations $\Pi_1, \ldots, \Pi_k$, each with unknown mean μ_i and known or unknown variance σ_i^2. Further denote the vector of means by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$, the vector of variances by $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_k^2)$, and the ordered μ_i's by $\mu_{[1]} \leq \cdots \leq \mu_{[k]}$. Our goal is to select the population associated with mean $\mu_{[k]}$. The population having mean $\mu_{[k]}$ is considered the "best" alternative, and a *correct selection* (CS) is said to be made if this goal is achieved.

The indifference-zone approach is based upon the probability requirement that for specified constants $(\delta^\star, P^\star)$ with $\delta^\star > 0$ and $1/k < P^\star < 1$, we require that

$$P\{\text{CS}\} \;\geq\; P^\star \quad \text{whenever} \quad \mu_{[k]} - \mu_{[k-1]} \;\geq\; \delta^\star. \tag{39}$$

This probability depends on the differences $\mu_i - \mu_j$ $(i \neq j,\ 1 \leq i, j \leq k)$, the sample size n, and $\boldsymbol{\sigma}^2$. The constant $\delta^\star$ can be thought of as the "smallest difference worth detecting," that is, any differences smaller than that are regarded as practically insignificant and are of no concern to us. Further define the parameter configurations $\boldsymbol{\mu}$ satisfying $\mu_{[k]} - \mu_{[k-1]} \geq \delta^\star$ as being in the *preference-zone*, $\Omega \equiv \{\boldsymbol{\mu} | \mu_{[k]} - \mu_{[k-1]} \geq \delta^\star\}$, for a correct selection and the configurations satisfying $\mu_{[k]} - \mu_{[k-1]} < \delta^\star$ as being in the *indifference-zone*, Ω^c. Any selection procedure that guarantees (39) is said to be employing the *indifference-zone* approach.

There are hundreds of procedures which address this type of problem, for example,

- Single-Stage Procedure (Bechhofer [1])
- Two-Stage Procedure (Rinott [15])
- Unbounded Sequential Procedure (Bechhofer, Kiefer, and Sobel [2])
- "Bounded" Sequential Procedure (Kim and Nelson [11])

Ranking and selection procedures such as those listed above are typically designed to obtain an a priori $P\{\text{CS}\}$ that is specified before the start of any experimentation — the $P\{\text{CS}\}$ requirement is one that will average out over the totality of observation realizations. This chapter of the thesis studies how the probability of correct selection is affected by the circumstances of procedure termination.

In parallel to addressing such conditional $P\{\text{CS}\}$ properties, we present details on a few of the more-popular procedures.

4.2 Single-Stage Procedure $\mathcal{N}_{\mathrm{B}}$

This fundamental procedure due to Bechhofer [1] assumes that the populations have *common known variance*, and for the given k and specified $(\delta^\star/\sigma, P^\star)$ determines a fixed sample size n (usually from a table — see below). We then take a random sample of n observations Y_{ij} $(1 \le j \le n)$ in a single stage from Π_i $(1 \le i \le k)$ and calculate the k sample means $\bar{Y}_i = \sum_{j=1}^{n} Y_{ij}/n$ $(1 \le i \le k)$. We select the population that yields the largest sample mean as the one associated with $\mu_{[k]}$.

One attractive feature of this procedure is how intuitive it is. Once we obtain the appropriate value of n, the procedure is straightforward. To choose n, we can either use a table such as Table 14, abstracted from Bechhofer, Santner, and Goldsman [3], or the formula

$$n = \left\lceil 2\left(\sigma Z^{(1-P^\star)}_{k-1,1/2}/\delta^\star\right)^2 \right\rceil,$$

where $Z^{(1-P^\star)}_{k-1,1/2}$ is a special case of the upper-equicoordinate point of a certain multivariate normal distribution. The constant Z is determined so as to satisfy the

Table 14: Common sample size n per population required by $\mathcal{N}_{\mathrm{B}}$

k	$P^\star$	$\delta^\star/\sigma$									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
2	0.75	91	23	11	6	4	3	2	2	2	1
	0.90	329	83	37	21	14	10	7	6	5	4
	0.95	542	136	61	34	22	16	12	9	7	6
	0.99	1083	271	121	68	44	31	23	17	14	11
3	0.75	206	52	23	13	9	6	5	4	3	3
	0.90	498	125	56	32	20	14	11	8	7	5
	0.95	735	184	82	46	30	21	15	12	10	8
	0.99	1309	328	146	82	53	37	27	21	17	14
4	0.75	283	71	32	18	12	8	6	5	4	3
	0.90	602	151	67	38	25	17	13	10	8	7
	0.95	851	213	95	54	35	24	18	14	11	9
	0.99	1442	361	161	91	58	41	30	23	18	15

probability requirement (39) for any true configuration of means $\boldsymbol{\mu} \in \Omega$. The configuration $\boldsymbol{\mu}$ in the preference zone Ω for which the minimum $P\{\mathrm{CS}\}$ is achieved is called the least-favorable (LF) configuration. It is often the case that the minimum $P\{\mathrm{CS}\}$ is achieved by the so-called slippage configuration (SC),

$$\mu_{[1]} \;=\; \mu_{[k-1]} \;=\; \mu_{[k]} - \delta^\star. \tag{40}$$

Example 12 Suppose that $k = 2$ and that we want to detect a difference in means as small as $\delta^\star/\sigma = 0.1$ standard deviations with $P\{\mathrm{CS}\}$ of at least 0.90. Procedure $\mathcal{N}_{\mathrm{B}}$ calls for $n = 329$ observations per population. $\square$

Another feature of this procedure is that it is relatively easy to obtain conditional $P\{\mathrm{CS}\}$ values (owing to the fact that the sample size n is fixed before the start of experimentation). For example, for $k = 2$ with $\sigma = 1$, suppose that $\mu_2 > \mu_1$, so that a CS corresponds to the selection of, say, population 2. Then

$$
\begin{aligned}
&P\Big\{\mathrm{CS}\,\Big|\,\max_i \bar{Y}_i = a, \min_i \bar{Y}_i = b\Big\} \\
&= P\Big\{\bar{Y}_1 = b,\, \bar{Y}_2 = a\,\Big|\Big(\bar{Y}_1 = b,\, \bar{Y}_2 = a\Big) \text{ or } \Big(\bar{Y}_1 = a,\, \bar{Y}_2 = b\Big)\Big\} \\
&= \frac{\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(b-\mu_1)^2}{2/\sqrt{n}}\}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(a-\mu_2)^2}{2/\sqrt{n}}\}}{\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(b-\mu_1)^2}{2/\sqrt{n}}\}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(a-\mu_2)^2}{2/\sqrt{n}}\} + \frac{1}{\sqrt{2\pi}}\exp\{-\frac{(a-\mu_1)^2}{2/\sqrt{n}}\}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{(b-\mu_2)^2}{2/\sqrt{n}}\}} \\
&= \Big[1 + \exp\{-(\mu_2 - \mu_1)(a-b)\sqrt{n}\}\Big]^{-1}.
\end{aligned}
$$

This can be readily generalized to other k and arbitrary (but known) variances.

Figure 14 plots $P\{\mathrm{CS}\}$ as a function of $a-b$. As the difference becomes large, $P\{\mathrm{CS}\}$ approaches 1. This intuitively makes sense, for if a population mean is significantly larger than another, it should be much easier for the experimenter to detect that difference and make the correct decision.

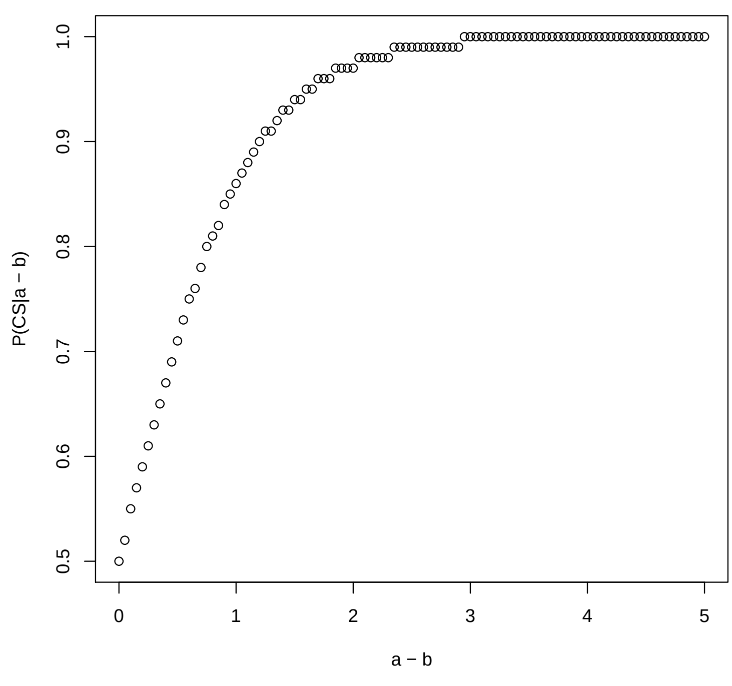

Figure 14: Conditional $P\{\mathrm{CS}\}$ of Single-Stage Procedure $\mathcal{N}_{\mathrm{B}}$.

Example 13 How close is our chosen value of $\delta^\star$ to the true difference between population means, and how does this affect the conditional probability of correct selection? To examine these effects, we calculated the probability of correct selection given that the procedure has stopped at a certain stage of sampling. In this example, we take $k = 3$, $\delta^\star = 0.1$, $P^\star = 0.9$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0.025$, and we let the true difference between population means range from 0.10 to 0.20. See Figure 15, which plots the conditional $P\{\text{CS}\}$ vs. the termination stage and the true difference in means (where $\mu_{[3]} - \mu_{[2]} = \mu_{[3]} - \mu_{[1]}$). Note that for ease of presentation, we only illustrate the plot for termination stages 15–20. Figure 15 shows that even for the more-conservative cases in which $\delta^\star$ is much smaller than the true difference between means, we can observe dips below the overall desired probability of correct selection $P^\star$. $\square$

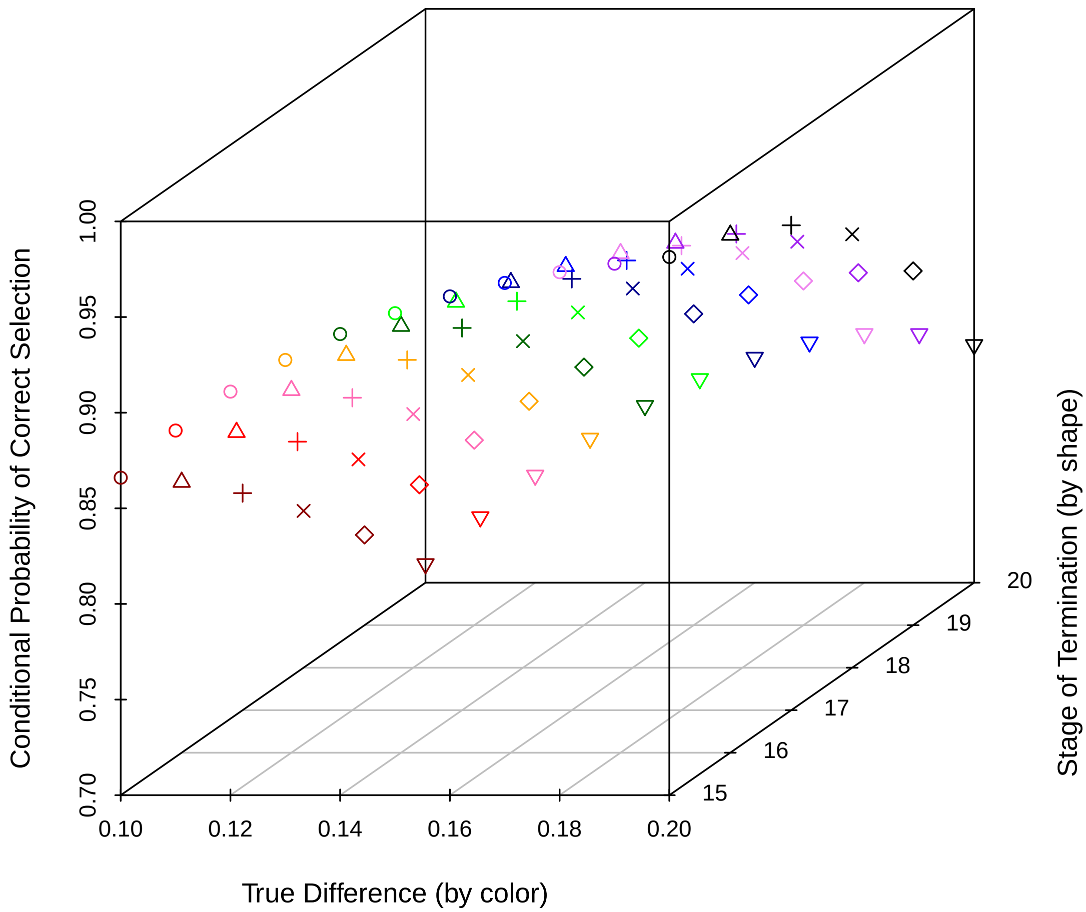

Figure 15: Effects of true difference between population means and stage of termination on conditional $P\{\text{CS}\}$.

4.3 Open Sequential Procedure $\mathcal{N}_{\text{BKS}}$

We now consider a sequential procedure due to Bechhofer, Kiefer, and Sobel (BKS) [2]. This procedure uses the same indifference-zone approach for selection as the fixed sample-size Procedure $\mathcal{N}_{\text{B}}$.

For given k, and common known σ, specify $P^\star$ and $\delta^\star$. At stage m of experimentation ($m \geq 1$), observe the random vector $(Y_{1m}, \ldots, Y_{km})$ and calculate $x_{im} = \sum_{j=1}^{m} y_{ij}$ ($1 \leq i \leq k$). Denote the ordered x_{im}'s by $x_{[1]m} < \cdots < x_{[k]m}$. Stop sampling when, for the first time,

$$z_m \equiv \sum_{i=1}^{k-1} \exp\{-\delta^\star(x_{[k]m} - x_{[i]m})/\sigma^2\} \leq \frac{(1 - P^\star)}{P^\star}.$$

Let N (a random variable) be the value of m when sampling stops. Select the treatment that yielded $x_{[k]N}$ as the one associated with $\mu_{[k]}$. Note that, although this is an unbounded open procedure, it is actually guaranteed to stop with probability one; and the procedure is in some sense similar to a sequential probability ratio test. Unfortunately, this procedure tends to be conservative in that it often delivers substantially higher $P\{\text{CS}\}$ than the desired $P^\star$, at the cost of larger $\text{E}[N]$. However, this conservatism can be mitigated by truncation—that is, stop sampling after a certain point that is chosen to guarantee the probability requirement (39).

BKS show that if $\boldsymbol{\mu}$ happens to be in the slippage configuration given by (40), then $W_N \equiv 1/(1 + z_N)$ is an unbiased estimator of the $P\{\text{CS}\}$. In fact, this result holds for a wide variety of distributions using a more-general version of the BKS procedure.

4.4 Closed Sequential Procedure $\mathcal{N}_{\text{KN}}$

Often, we do not know the true values of the population variances and cannot always ensure their equality. The procedure of Kim and Nelson (KN) [11] assumes that the populations have unknown (and possibly unequal) variances. The KN procedure efficiently eliminates populations that it deems as inferior until only one population remains.

It proceeds as follows. For the given k, specify $(\delta^\star, P^\star)$, and a common initial sample size $n_0 \geq 2$ to be taken from each scenario. Calculate the constant (which will be used later)

$$\eta = \frac{1}{2}\left[\left(\frac{2(1-P^\star)}{k-1}\right)^{-2/(n_0-1)} - 1\right].$$

Set the initial set of retained populations $I = \{1, 2, \ldots, k\}$ and let $h = \sqrt{2\eta(n_0 - 1)}$.

Take a random sample of n_0 observations Y_{ij} $(1 \le j \le n_0)$ from population i $(1 \le i \le k)$. For population i, compute the sample mean based on the n_0 observations, $\bar{Y}_i(n_0) = \sum_{j=1}^{n_0} Y_{ij}/n_0$ $(1 \le i \le k)$. For all $i \neq \ell$, compute the sample variance of the difference between treatments i and ℓ,

$$S_{i\ell}^2 = \frac{1}{n_0 - 1}\sum_{j=1}^{n_0}\Big(Y_{ij} - Y_{\ell j} - [\bar{Y}_i(n_0) - \bar{Y}_\ell(n_0)]\Big)^2.$$

Further, for all $i \neq \ell$, set

$$N_{i\ell} = \left\lfloor h^2 S_{i\ell}^2/(\delta^\star)^2 \right\rfloor,$$

where $\lfloor\cdot\rfloor$ is the floor function, and for all i, set

$$N_i = \max_{\ell \neq i} N_{i\ell}.$$

If $n_0 > \max_i N_i$, stop and select the population with the largest sample mean $\bar{Y}_i(n_0)$ as one having the largest mean. Otherwise, set the sequential counter $r = n_0$ and go to the Screening phase of the procedure.

Screening: Set $I^{\text{old}} = I$ and revise the set of retained populations,

$$I = \{i : i \in I^{\text{old}} \text{ and } \bar{Y}_i(r) \ge \bar{Y}_\ell(r) - W_{i\ell}(r), \text{ for all } \ell \in I^{\text{old}}, \ell \neq i\},$$

where the "whisker length"

$$W_{i\ell}(r) = \max\left\{0, \frac{\delta^\star}{2r}\left(\frac{h^2 S_{i\ell}^2}{(\delta^\star)^2} - r\right)\right\}.$$

In other words, keep in play those surviving populations that are not "too far" from the current leader (and eliminate the others).

Stopping Rule: If $|I| = 1$, then stop and select the population with index in I as having the largest mean.

If $|I| > 1$, take one additional observation $Y_{i,r+1}$ from each treatment $i \in I$. Increment $r = r + 1$ and go to the Screening stage if $r < \max_i N_i + 1$. On the other hand, if $r = \max_i N_i + 1$, then stop and select the treatment associated with the largest $\bar{Y}_i(r)$ having index $i \in I$.

The question of the hour is: Is the $P\{\text{CS}\}$ of the KN procedure affected by when and where the procedure stops? Yes. The procedure is designed to give an overall $P\{\text{CS}\} \geq P^\star$ for $\boldsymbol{\mu} \in \Omega$. But the conditional $P\{\text{CS}\}$ given a stop at a certain stage can differ substantially from the overall $P\{\text{CS}\}$ both on the high and low sides. Of course, we can use simulation to find this conditional $P\{\text{CS}\}$ for given mean- and variance-vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$, but some of the stopping points for Procedure $\mathcal{N}_{\text{KN}}$ occur with very low probability, often making simulation of these rare events problematic.

We now discuss results from Procedure $\mathcal{P}$ of Wang and Kim [18], which is a special case of Procedure $\mathcal{N}_{\text{KN}}$ and is designed for the known (but not necessarily common) variance case. Procedure $\mathcal{P}$ is almost identical to Procedure $\mathcal{N}_{\text{KN}}$, except that all variance estimates are essentially replaced by their exact values. We will be able to undertake certain exact calculations for the Wang and Kim procedure; we run Procedure $\mathcal{P}$ as follows.

Suppose $Y_{ir} \sim \text{Nor}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, k$, $r \geq 1$, and let

$$X_{ij}^{(\ell)} \ = \ \sum_{r=1}^{\ell} (Y_{ir} - Y_{jr}), \quad \ell \geq 1,$$

which compares the sample sums of populations i and j at the ℓth stage. Clearly,

$$\begin{bmatrix} X_{12}^{(1)} \\ \vdots \\ X_{(k-1)k}^{(1)} \\ X_{12}^{(2)} \\ \vdots \\ X_{(k-1)k}^{(2)} \\ \vdots \\ X_{(k-1)k}^{(n_{\max})} \end{bmatrix} \sim \text{ multivariate normal,}$$

where $n_{\max} = \max(N_i)$ is the maximum possible number of observations taken, and

$$\begin{aligned}
\mathrm{E}[X_{ij}^{(\ell)}] &= \ell(\mu_i - \mu_j) \\
\mathrm{Var}(X_{ij}^{(\ell)}) &= \ell(\sigma_i^2 + \sigma_j^2) \\
\mathrm{Cov}(X_{ij}^{(\ell)}, X_{ij}^{(m)}) &= \min(\ell, m)(\sigma_i^2 + \sigma_j^2) \\
\mathrm{Cov}(X_{ij}^{(m)}, X_{i\ell}^{(n)}) &= \min(m, n)\sigma_i^2 \\
\mathrm{Cov}(X_{ij}^{(m)}, X_{\ell j}^{(n)}) &= \min(m, n)\sigma_j^2 \\
\mathrm{Cov}(X_{ij}^{(m)}, X_{j\ell}^{(n)}) &= \mathrm{Cov}(X_{ji}^{(m)}, X_{\ell j}^{(n)}) \;=\; -\min(m, n)\sigma_j^2 \\
\mathrm{Cov}(X_{ij}^{(m)}, X_{ji}^{(n)}) &= -\min(m, n)(\sigma_i^2 + \sigma_j^2)
\end{aligned}$$

We can use readily available software (say, in R) for calculating multivariate normal probabilities; and by keeping track of all the ways that a procedure can stop at some stage i, we can calculate all of the stopping probabilities and CPCS results for this procedure.

Example 14 We apply the Wang and Kim procedure to the problem $k = 3$, $P^\star = 0.90$, $\delta^\star = 0.1$, $\boldsymbol{\mu} = (0.1, 0, 0)$, and known $\boldsymbol{\sigma} = (0.015, 0.015, 0.015)$. We obtain exact probabilities of procedure termination at each stage, the conditional $P\{\mathrm{CS}\}$ given termination at each stage, and the overall $P\{\mathrm{CS}\} = 0.936$. See Figure 16, which shows that, in this case, the CPCS tends to decrease as the termination stage increases. $\square$

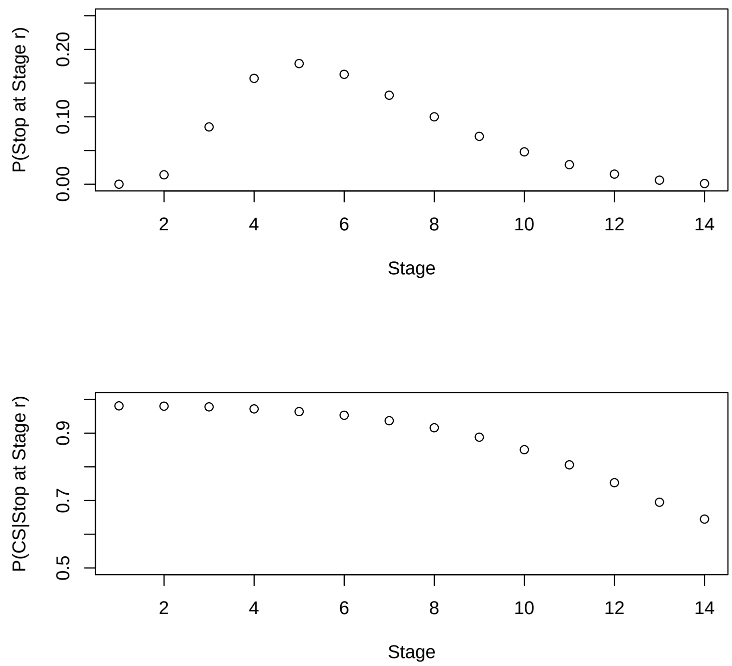

Figure 16: Wang and Kim Procedure $\mathcal{P}$ for the known variance case

For the unknown variance case, the exact calculations can be complex enough to merit the use of simulation to estimate the overall $P\{\text{CS}\}$ and the conditional $P\{\text{CS}\}$ at each stage.

Example 15 We apply Procedure $\mathcal{N}_{\text{KN}}$ to the scenario $k = 3$, $P^\star = 0.90$, $\delta^\star = 0.1$, $\boldsymbol{\mu} = (0.1, 0, 0)$, and (unknown) $\boldsymbol{\sigma} = (0.01, 0.01, 0.01)$; the initial sample size is $n_0 = 20$. We use Monte Carlo simulation (1,000,000 replications) to estimate $P\{\text{CS}\} \approx 0.925$, as well as the conditional $P\{\text{CS}\}$ at each stage. See Figure 17, which shows that, as in Example 14, the CPCS tends to decrease as the termination stage increases. $\square$

Example 16 We conduct additional experimentation to further study the effects of how and when we stop on the CPCS. The experiments are based on 10^6 replications of Wang and Kim's Procedure $\mathcal{P}$ for $P^\star = 0.90$ and various levels of $\delta^\star$ and variance configurations $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. Table 15 concerns the case in which the $k = 3$ populations

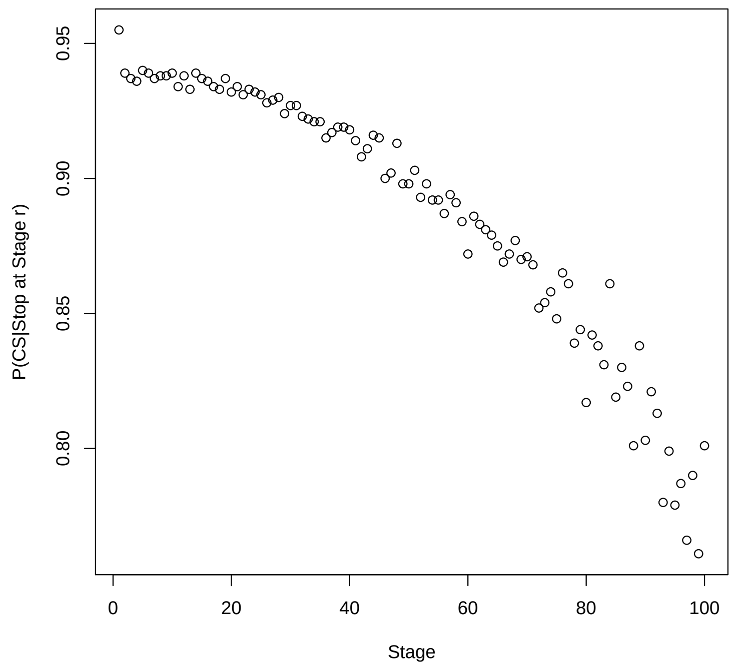

Figure 17: $\mathcal{N}_{\mathrm{KN}}$ for the unknown variance case

all have the same variance; Table 16 is for the case when the best population's variance is greater than those of the other two; and Table 17 is for the case when the best population's variance is smaller than those of the other two. Generally speaking, for fixed $\boldsymbol{\sigma}$, the procedure is more likely to terminate early as $\delta^\star$ increases; and in the cases in which the procedure does take multiple stages to terminate, the CPCS drops well below the guaranteed overall probability of correct selection 0.90 guaranteed by the procedure. $\square$

Example 17 We also conducted a robustness study in which we used observations from exponential and Pareto distributions instead of the normal distribution. These results can be found in Tables 18 and 19, respectively. The scenario is analogous to the setup for Table 15, and the results are qualitatively similar. The overall $P\{\mathrm{CS}\}$ for the procedure is still well over $P^\star = 0.90$, so the procedure is still valid. These examples further demonstrate how $\delta^\star$ and stage of the procedure affect the CPCS, as well as illustrate that these procedures are fairly robust, and that a (hopefully minor) error in initial assumptions may not result in an invalid selection. □

Table 15: CPCS of Wang and Kim's Procedure $\mathcal{P}$ using $k = 3$ populations, $P^\star = 0.9$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$, 10^6 replications, and the slippage configuration.

	$\delta^\star = 0.5$		$\delta^\star = 0.6$		$\delta^\star = 0.7$		$\delta^\star = 0.8$		$\delta^\star = 0.9$		$\delta^\star = 1.0$	
Stage	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS
1	0.0000		0.0000		0.0000		0.0000	1.0000	0.0004	0.9890	0.0021	0.9833
2	0.0000	1.0000	0.0001	0.9911	0.0018	0.9761	0.0113	0.9812	0.0391	0.9796	0.0947	0.9795
3	0.0002	0.9935	0.0037	0.9785	0.0229	0.9818	0.0738	0.9784	0.1521	0.9755	0.2395	0.9732
4	0.0020	0.9769	0.0193	0.9792	0.0695	0.9770	0.1440	0.9729	0.2121	0.9682	0.2493	0.9605
5	0.0083	0.9767	0.0452	0.9756	0.1112	0.9717	0.1717	0.9661	0.1990	0.9559	0.1865	0.9396
6	0.0195	0.9774	0.0714	0.9732	0.1316	0.9663	0.1615	0.9552	0.1537	0.9373	0.1191	0.9067
7	0.0330	0.9753	0.0899	0.9687	0.1325	0.9584	0.1351	0.9405	0.1076	0.9094	0.0678	0.8595
8	0.0462	0.9717	0.0988	0.9642	0.1202	0.9487	0.1052	0.9209	0.0697	0.8737	0.0313	0.7899
9	0.0572	0.9699	0.0993	0.9573	0.1036	0.9377	0.0771	0.8980	0.0404	0.8245	0.0092	0.7158
10	0.0646	0.9664	0.0937	0.9517	0.0844	0.9217	0.0536	0.8644	0.0194	0.7629	0.0004	0.6847
11	0.0693	0.9635	0.0863	0.9428	0.0676	0.9056	0.0347	0.8257	0.0060	0.7032	0.0000	
12	0.0701	0.9590	0.0765	0.9344	0.0521	0.8795	0.0197	0.7798	0.0005	0.6559	0.0000	
13	0.0700	0.9546	0.0660	0.9233	0.0389	0.8522	0.0092	0.7246	0.0000		0.0000	
14	0.0670	0.9490	0.0562	0.9098	0.0276	0.8244	0.0027	0.6619	0.0000		0.0000	
15	0.0629	0.9427	0.0472	0.8946	0.0183	0.7814	0.0002	0.6222	0.0000		0.0000	
16	0.0577	0.9370	0.0386	0.8762	0.0108	0.7434	0.0000		0.0000		0.0000	
17	0.0533	0.9311	0.0313	0.8572	0.0052	0.6900	0.0000		0.0000		0.0000	
18	0.0477	0.9212	0.0247	0.8314	0.0018	0.6558	0.0000		0.0000		0.0000	
19	0.0430	0.9105	0.0184	0.8021	0.0003	0.5745	0.0000		0.0000		0.0000	
20	0.0382	0.9030	0.0136	0.7746	0.0000		0.0000		0.0000		0.0000	
21	0.0335	0.8905	0.0092	0.7471	0.0000		0.0000		0.0000		0.0000	
22	0.0296	0.8780	0.0057	0.7123	0.0000		0.0000		0.0000		0.0000	
23	0.0257	0.8658	0.0031	0.6678	0.0000		0.0000		0.0000		0.0000	
24	0.0220	0.8473	0.0013	0.6512	0.0000		0.0000		0.0000		0.0000	
25	0.0183	0.8324	0.0004	0.6068	0.0000		0.0000		0.0000		0.0000	
26	0.0156	0.8145	0.0000	0.6818	0.0000		0.0000		0.0000		0.0000	
27	0.0126	0.7961	0.0000		0.0000		0.0000		0.0000		0.0000	
28	0.0098	0.7797	0.0000		0.0000		0.0000		0.0000		0.0000	
29	0.0079	0.7454	0.0000		0.0000		0.0000		0.0000		0.0000	
30	0.0148	0.6145	0.0000		0.0000		0.0000		0.0000		0.0000	
$P\{\text{CS}\}$	0.9266		0.9306		0.9333		0.9358		0.9381		0.9405	

Table 16: CPCS of Wang and Kim's Procedure $\mathcal{P}$ using $k = 3$ populations, $P^\star = 0.9$, $\sigma_1^2 = 1$, $\sigma_2^2 = \sigma_3^2 = 0.5$, 10^6 replications, and the slippage configuration.

	$\delta^\star = 0.5$		$\delta^\star = 0.6$		$\delta^\star = 0.7$		$\delta^\star = 0.8$		$\delta^\star = 0.9$		$\delta^\star = 1.0$	
Stage	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS
1	0.0000		0.0000	0.9651	0.0001	0.9587	0.0012	0.9605	0.0063	0.9616	0.0215	0.9603
2	0.0001	0.9616	0.0028	0.9610	0.0193	0.9601	0.0662	0.9586	0.1508	0.9559	0.2621	0.9528
3	0.0037	0.9606	0.0295	0.9578	0.0973	0.9534	0.1950	0.9476	0.2865	0.9391	0.3419	0.9273
4	0.0186	0.9566	0.0787	0.9506	0.1661	0.9416	0.2349	0.9285	0.2580	0.9091	0.2366	0.8827
5	0.0431	0.9509	0.1189	0.9411	0.1839	0.9251	0.2027	0.9007	0.1773	0.8677	0.1256	0.8336
6	0.0676	0.9446	0.1368	0.9290	0.1671	0.9036	0.1501	0.8663	0.1034	0.8271	0.0517	0.7624
7	0.0857	0.9370	0.1357	0.9144	0.1368	0.8773	0.1001	0.8320	0.0507	0.7696	0.0122	0.6637
8	0.0955	0.9281	0.1238	0.8967	0.1046	0.8494	0.0597	0.7980	0.0184	0.6799	0.0000	
9	0.0980	0.9180	0.1070	0.8765	0.0752	0.8237	0.0314	0.7345	0.0023	0.6238	0.0000	
10	0.0951	0.9064	0.0888	0.8552	0.0503	0.7977	0.0125	0.6571	0.0000		0.0000	
11	0.0889	0.8936	0.0713	0.8348	0.0314	0.7510	0.0024	0.5927	0.0000		0.0000	
12	0.0810	0.8792	0.0551	0.8175	0.0173	0.6914	0.0000		0.0000		0.0000	
13	0.0721	0.8641	0.0410	0.7978	0.0074	0.6271	0.0000		0.0000		0.0000	
14	0.0630	0.8492	0.0294	0.7665	0.0018	0.5664	0.0000		0.0000		0.0000	
15	0.0541	0.8346	0.0200	0.7250	0.0000	0.5568	0.0000		0.0000		0.0000	
16	0.0457	0.8216	0.0125	0.6780	0.0000		0.0000		0.0000		0.0000	
17	0.0378	0.8093	0.0067	0.6276	0.0000		0.0000		0.0000		0.0000	
18	0.0306	0.7962	0.0027	0.5821	0.0000		0.0000		0.0000		0.0000	
19	0.0244	0.7769	0.0006	0.5364	0.0000		0.0000		0.0000		0.0000	
20	0.0191	0.7499	0.0000	0.5345	0.0000		0.0000		0.0000		0.0000	
21	0.0145	0.7200	0.0000		0.0000		0.0000		0.0000		0.0000	
22	0.0105	0.6865	0.0000		0.0000		0.0000		0.0000		0.0000	
23	0.0071	0.6525	0.0000		0.0000		0.0000		0.0000		0.0000	
24	0.0044	0.6138	0.0000		0.0000		0.0000		0.0000		0.0000	
25	0.0024	0.5782	0.0000		0.0000		0.0000		0.0000		0.0000	
26	0.0010	0.5465	0.0000		0.0000		0.0000		0.0000		0.0000	
27	0.0003	0.5115	0.0000		0.0000		0.0000		0.0000		0.0000	
28	0.0000	0.4831	0.0000		0.0000		0.0000		0.0000		0.0000	
29	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	
30	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	
$P\{\text{CS}\}$	0.9358		0.9386		0.9412		0.9438		0.9462		0.9485	

Table 17: CPCS of Wang and Kim's Procedure $\mathcal{P}$ using $k = 3$ populations, $P^\star = 0.9$, $\sigma_1^2 = 0.5$, $\sigma_2^2 = \sigma_3^2 = 1$, 10^6 replications, and the slippage configuration.

	$\delta^\star = 0.5$		$\delta^\star = 0.6$		$\delta^\star = 0.7$		$\delta^\star = 0.8$		$\delta^\star = 0.9$		$\delta^\star = 1.0$	
Stage	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS
1	0.0000		0.0000	1.0000	0.0000	0.9671	0.0003	0.9636	0.0021	0.9627	0.0095	0.9623
2	0.0000	0.9614	0.0008	0.9627	0.0084	0.9629	0.0383	0.9618	0.1058	0.9602	0.2105	0.9580
3	0.0011	0.9636	0.0149	0.9619	0.0667	0.9594	0.1616	0.9549	0.2682	0.9481	0.3459	0.9381
4	0.0089	0.9619	0.0549	0.9575	0.1436	0.9510	0.2301	0.9398	0.2709	0.9223	0.2583	0.8954
5	0.0271	0.9585	0.0997	0.9506	0.1797	0.9374	0.2140	0.9154	0.1951	0.8789	0.1452	0.8198
6	0.0508	0.9537	0.1281	0.9409	0.1741	0.9186	0.1644	0.8793	0.1199	0.8114	0.0659	0.6979
7	0.0723	0.9478	0.1359	0.9284	0.1475	0.8927	0.1138	0.8275	0.0639	0.7108	0.0190	0.5041
8	0.0874	0.9401	0.1292	0.9124	0.1157	0.8586	0.0725	0.7561	0.0262	0.5665	0.0030	0.0000
9	0.0949	0.9321	0.1146	0.8929	0.0860	0.8146	0.0409	0.6610	0.0062	0.2824	0.0012	0.0000
10	0.0960	0.9218	0.0971	0.8688	0.0608	0.7586	0.0186	0.5275	0.0022	0.0000	0.0001	0.0000
11	0.0923	0.9101	0.0796	0.8397	0.0401	0.6870	0.0058	0.2897	0.0009	0.0000	0.0000	
12	0.0857	0.8964	0.0635	0.8038	0.0238	0.5972	0.0021	0.0000	0.0001	0.0000	0.0000	
13	0.0776	0.8806	0.0491	0.7617	0.0118	0.4729	0.0012	0.0000	0.0000		0.0000	
14	0.0687	0.8625	0.0368	0.7102	0.0047	0.2681	0.0005	0.0000	0.0000		0.0000	
15	0.0601	0.8417	0.0262	0.6512	0.0020	0.0121	0.0000	0.0000	0.0000		0.0000	
16	0.0516	0.8185	0.0174	0.5786	0.0014	0.0000	0.0000		0.0000		0.0000	
17	0.0439	0.7912	0.0103	0.4851	0.0008	0.0000	0.0000		0.0000		0.0000	
18	0.0367	0.7606	0.0054	0.3536	0.0003	0.0000	0.0000		0.0000		0.0000	
19	0.0301	0.7264	0.0026	0.1607	0.0001	0.0000	0.0000		0.0000		0.0000	
20	0.0243	0.6882	0.0015	0.0085	0.0000		0.0000		0.0000		0.0000	
21	0.0191	0.6426	0.0012	0.0000	0.0000		0.0000		0.0000		0.0000	
22	0.0144	0.5933	0.0008	0.0000	0.0000		0.0000		0.0000		0.0000	
23	0.0104	0.5348	0.0005	0.0000	0.0000		0.0000		0.0000		0.0000	
24	0.0070	0.4634	0.0002	0.0000	0.0000		0.0000		0.0000		0.0000	
25	0.0045	0.3684	0.0001	0.0000	0.0000		0.0000		0.0000		0.0000	
26	0.0027	0.2428	0.0000	0.0000	0.0000		0.0000		0.0000		0.0000	
27	0.0016	0.1002	0.0000		0.0000		0.0000		0.0000		0.0000	
28	0.0012	0.0141	0.0000		0.0000		0.0000		0.0000		0.0000	
29	0.0010	0.0000	0.0000		0.0000		0.0000		0.0000		0.0000	
30	0.0023	0.0000	0.0000		0.0000		0.0000		0.0000		0.0000	
$P\{\text{CS}\}$	0.9263		0.9296		0.9328		0.9357		0.9386		0.9413	

Table 18: CPCS of Wang and Kim's Procedure $\mathcal{P}$ using $k = 3$ populations, $P^\star = 0.9$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$, 10^6 replications, and the slippage configuration. Here we use exponential observations instead of normals.

	$\delta^\star = 0.5$		$\delta^\star = 0.6$		$\delta^\star = 0.7$		$\delta^\star = 0.8$		$\delta^\star = 0.9$		$\delta^\star = 1.0$	
Stage	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS
1	0.0001	0.5200	0.0006	0.5764	0.0022	0.6203	0.0053	0.6585	0.0111	0.6988	0.0202	0.7363
2	0.0007	0.6917	0.0037	0.7478	0.0114	0.8114	0.0278	0.8541	0.0572	0.8861	0.1054	0.9108
3	0.0027	0.8074	0.0120	0.8645	0.0349	0.8975	0.0786	0.9240	0.1465	0.9412	0.2294	0.9504
4	0.0068	0.8767	0.0271	0.9129	0.0709	0.9341	0.1379	0.9467	0.2087	0.9551	0.2546	0.9556
5	0.0142	0.9038	0.0484	0.9332	0.1068	0.9468	0.1680	0.9529	0.2002	0.9528	0.1867	0.9445
6	0.0236	0.9258	0.0693	0.9464	0.1269	0.9528	0.1627	0.9494	0.1537	0.9420	0.1128	0.9266
7	0.0346	0.9383	0.0861	0.9494	0.1301	0.9512	0.1353	0.9436	0.1038	0.9264	0.0595	0.9007
8	0.0461	0.9467	0.0957	0.9518	0.1204	0.9473	0.1031	0.9331	0.0638	0.9075	0.0247	0.8670
9	0.0551	0.9488	0.0971	0.9513	0.1032	0.9400	0.0738	0.9167	0.0348	0.8821	0.0065	0.8110
10	0.0622	0.9506	0.0934	0.9463	0.0840	0.9314	0.0498	0.8993	0.0154	0.8489	0.0003	0.7695
11	0.0668	0.9507	0.0855	0.9424	0.0661	0.9186	0.0309	0.8787	0.0046	0.7959	0.0000	
12	0.0685	0.9494	0.0758	0.9359	0.0496	0.9044	0.0167	0.8502	0.0003	0.7559	0.0000	
13	0.0677	0.9483	0.0659	0.9288	0.0362	0.8864	0.0077	0.8122	0.0000		0.0000	
14	0.0659	0.9448	0.0558	0.9222	0.0250	0.8725	0.0023	0.7686	0.0000		0.0000	
15	0.0620	0.9403	0.0460	0.9126	0.0163	0.8410	0.0002	0.7182	0.0000		0.0000	
16	0.0572	0.9389	0.0376	0.8971	0.0095	0.8216	0.0000		0.0000		0.0000	
17	0.0529	0.9314	0.0297	0.8863	0.0046	0.7811	0.0000		0.0000		0.0000	
18	0.0478	0.9253	0.0228	0.8710	0.0016	0.7590	0.0000		0.0000		0.0000	
19	0.0426	0.9206	0.0173	0.8532	0.0003	0.6702	0.0000		0.0000		0.0000	
20	0.0381	0.9099	0.0124	0.8423	0.0000		0.0000		0.0000		0.0000	
21	0.0337	0.9061	0.0082	0.8213	0.0000		0.0000		0.0000		0.0000	
22	0.0290	0.8935	0.0051	0.7977	0.0000		0.0000		0.0000		0.0000	
23	0.0249	0.8864	0.0029	0.7653	0.0000		0.0000		0.0000		0.0000	
24	0.0213	0.8757	0.0012	0.7357	0.0000		0.0000		0.0000		0.0000	
25	0.0177	0.8665	0.0004	0.6619	0.0000		0.0000		0.0000		0.0000	
26	0.0147	0.8523	0.0000	0.6444	0.0000		0.0000		0.0000		0.0000	
27	0.0121	0.8439	0.0000		0.0000		0.0000		0.0000		0.0000	
28	0.0094	0.8261	0.0000		0.0000		0.0000		0.0000		0.0000	
29	0.0072	0.8038	0.0000		0.0000		0.0000		0.0000		0.0000	
30	0.0145	0.6760	0.0000		0.0000		0.0000		0.0000		0.0000	
$P\{\text{CS}\}$	0.9237		0.9280		0.9293		0.9307		0.9328		0.9335	

Table 19: CPCS of Wang and Kim's Procedure $\mathcal{P}$ using $k = 3$ populations, $P^\star = 0.9$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$, 10^6 replications, and the slippage configuration. Here we use Pareto observations instead of normals.

	$\delta^\star = 0.5$		$\delta^\star = 0.6$		$\delta^\star = 0.7$		$\delta^\star = 0.8$		$\delta^\star = 0.9$		$\delta^\star = 1.0$	
Stage	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS	P(Stop)	CPCS
1	0.0037	0.3867	0.0056	0.4170	0.0086	0.4497	0.0122	0.4722	0.0167	0.5196	0.0228	0.5607
2	0.0046	0.4885	0.0080	0.5589	0.0136	0.6260	0.0231	0.7147	0.0394	0.7921	0.0688	0.8623
3	0.0060	0.5801	0.0120	0.6960	0.0249	0.8021	0.0526	0.8757	0.1160	0.9336	0.2487	0.9649
4	0.0082	0.6858	0.0199	0.8126	0.0494	0.9024	0.1235	0.9510	0.2631	0.9745	0.3546	0.9822
5	0.0117	0.7751	0.0335	0.8901	0.0936	0.9486	0.2113	0.9744	0.2678	0.9818	0.1827	0.9794
6	0.0170	0.8469	0.0555	0.9316	0.1458	0.9703	0.2172	0.9805	0.1542	0.9776	0.0777	0.9706
7	0.0245	0.8930	0.0820	0.9581	0.1708	0.9774	0.1526	0.9786	0.0782	0.9718	0.0321	0.9613
8	0.0353	0.9262	0.1077	0.9686	0.1526	0.9791	0.0924	0.9752	0.0387	0.9638	0.0107	0.9469
9	0.0480	0.9462	0.1204	0.9753	0.1149	0.9761	0.0547	0.9686	0.0176	0.9518	0.0020	0.9084
10	0.0612	0.9582	0.1185	0.9769	0.0804	0.9733	0.0311	0.9610	0.0066	0.9267	0.0000	0.8605
11	0.0720	0.9654	0.1028	0.9768	0.0547	0.9688	0.0171	0.9519	0.0015	0.8878	0.0000	
12	0.0793	0.9708	0.0848	0.9746	0.0363	0.9638	0.0083	0.9429	0.0001	0.8286	0.0000	
13	0.0816	0.9733	0.0666	0.9710	0.0236	0.9556	0.0031	0.9220	0.0000		0.0000	
14	0.0796	0.9739	0.0510	0.9672	0.0149	0.9475	0.0007	0.8770	0.0000		0.0000	
15	0.0737	0.9742	0.0386	0.9626	0.0086	0.9388	0.0000	0.7391	0.0000		0.0000	
16	0.0666	0.9727	0.0288	0.9578	0.0045	0.9298	0.0000		0.0000		0.0000	
17	0.0579	0.9723	0.0211	0.9551	0.0020	0.9014	0.0000		0.0000		0.0000	
18	0.0501	0.9693	0.0153	0.9477	0.0006	0.8844	0.0000		0.0000		0.0000	
19	0.0425	0.9665	0.0109	0.9402	0.0001	0.6733	0.0000		0.0000		0.0000	
20	0.0351	0.9657	0.0073	0.9301	0.0000		0.0000		0.0000		0.0000	
21	0.0295	0.9588	0.0046	0.9176	0.0000		0.0000		0.0000		0.0000	
22	0.0240	0.9571	0.0028	0.9119	0.0000		0.0000		0.0000		0.0000	
23	0.0198	0.9529	0.0015	0.8966	0.0000		0.0000		0.0000		0.0000	
24	0.0163	0.9480	0.0007	0.8626	0.0000		0.0000		0.0000		0.0000	
25	0.0131	0.9429	0.0002	0.8543	0.0000		0.0000		0.0000		0.0000	
26	0.0104	0.9412	0.0000	0.7895	0.0000		0.0000		0.0000		0.0000	
27	0.0081	0.9304	0.0000		0.0000		0.0000		0.0000		0.0000	
28	0.0062	0.9284	0.0000		0.0000		0.0000		0.0000		0.0000	
29	0.0046	0.9173	0.0000		0.0000		0.0000		0.0000		0.0000	
30	0.0092	0.7816	0.0000		0.0000		0.0000		0.0000		0.0000	
$P\{$CS$\}$	0.9473		0.9512		0.9534		0.9546		0.9559		0.9574	

CHAPTER V

CONCLUSIONS AND FURTHER WORK

5.1 Conclusions

We have proposed a full quasi-Monte Carlo method based on antithetic variates and quasi-Monte Carlo sampling that will enable an experimenter to more accurately and precisely model the Asian averages and options. This method's computational formula enables the experimenter to produce an estimate for the expected value of an Asian average with a greatly reduced variance while only marginally increasing computational time. We have also put forth a compromise from the naive estimator and the full 2^m FQMC estimator that is especially useful in cases for which a transformation of the data is required or higher-order moments need to be estimated. This compromise partial QMC estimator allows us to obtain much of the variance reduction provided by the FQMC estimator while still keeping computational time down to reasonable levels.

We have applied these techniques along with the Gram–Charlier p.d.f. estimation methodology to more accurately estimate a target p.d.f. using a modest sample size. This allows one to quickly estimate the option price, quantiles, various probabilities involving an Asian average, and anything else that one could compute if the p.d.f. were known.

We have demonstrated that the post-hoc evaluation of ranking and selection methods could give an experimenter valuable insight into how strong of a conclusion the experimenter is in fact making. Using this type of analysis, one can better ascertain the strength of their results, whether the results are definitive, or if further analysis is required.

5.2 *Further Work*

Generally speaking, we will continue to work with the partial quasi-Monte Carlo PQMC Asian average estimator (a compromise between the FQMC estimator and the naive estimator), particularly in cases where the FQMC estimator is not feasible to use. A better understanding of this estimator's performance will allow us to choose an optimal balance between variance reduction and computational time.

In addition, it is of interest to incorporate other well-known variance reduction methods such as importance sampling, stratified sampling, control variates, etc. into our methodology; such variation reduction techniques could significantly enhance the efficacy of our procedures. We have made progress in reducing the physical time it takes to calculate some our estimators. For example, although not reported here, we can calculate the full permutation estimator for $\mathrm{E}[A_m^+]$ via a simple iterative method. The future holds a time study to compare the computational efficiencies of our various estimators.

In parallel with the above, we will apply these efficient estimators in conjunction with the Gram–Charlier method to estimate the probability density functions of additional classes of examples. The battery of examples will serve to establish how robust our procedures are.

We also continue our investigation into how the conditional probability of correct selection manifests in other ranking and selection procedures. Specifically, we will examine the conditional $P\{\mathrm{CS}\}$ performance on some of the more-recent procedures involving Bernoulli and multinomial sampling (for example, Tollefson et al. [16, 17]).

REFERENCES

[1] Bechofer, R. E. 1954. A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Annals of Matheatical Statistics*, 25:16–39.

[2] Bechhofer, R. E., Kiefer, J., and Sobel, M. 1968. *Sequential Identification and Ranking Procedures (with Special Reference to Koopman-Darmois Populations)*. Chicago: University of Chicago Press.

[3] Bechhofer, R. E., Santner, T. J., and Goldsman, D. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*. New York: John Wiley and Sons.

[4] Berberan-Santos, M. 2007. Expressing a probability density function in terms of another pdf: A generalized Gram–Charlier expansion. *Journal of Mathematical Chemistry*, 42:585–594.

[5] Billingsley, P. 1968. *Convergence of Probability Measures*. New York: John Wiley and Sons.

[6] Black, F., and Scholes, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.

[7] Carverhill, A., and Clewlow, L. 1990. Flexible convolution: Valuing average rate (Asian) options. *Risk*, 3(4):25–29.

[8] Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Heidelberg: Springer-Verlag.

[9] Gradshteyn, I. S., and Ryzhik, I. M. 1980. *Table of Integrals, Series, and Products*. New York: Academic Press.

[10] Kemna, A. G. Z., and Vorst, A. C. F. 1990. A pricing method for options based on average asset values. *Journal of Banking and Finance*, 14:113–129.

[11] Kim, S., and Nelson, B. L. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation*, 11:251–273.

[12] Law, A. M. 2013. *Simulation Modeling and Analysis* (5th edition). New York: McGraw-Hill Education.

[13] Popovic, R., and Goldsman, D. 2012. On valuing and hedging European options when volatility is estimated directly. *European Journal of Operational Research*, 218:124–131.

[14] Popovic, R., and Goldsman, D. 2012. Easy Gram–Charlier valuations of options. *Journal of Derivatives*, 20(2):79–97.

[15] Rinott, Y. 1978. On two-stage procedures and related probability-inequalities. *Communication in Statistics — Theory and Methods*, A8:799–811.

[16] Tollefson, E., Goldsman, D., Kleywegt, A., and Tovey, C., 2014. A Comparative Study of Procedures for the Multinomial Selection Problem. In *Essays in Production, Project Planning and Scheduling: A Festschrift in Honor of Salah Elmaghraby* (ed. P. S. Pulat, S. C. Sarin, and R. Uzsoy), 123–160, International Series in Operations Research & Management Science, Volume 200. New York: Springer.

[17] Tollefson, E., Goldsman, D., Kleywegt, A., and Tovey, C., 2014. Optimal Selection of the Most Probable Multinomial Alternative. *Sequential Analysis* 33: 491–508.

[18] Wang, H. and Kim, S.-H. 2013. Reducing the conservativeness of fully sequential indifference-zone procedures. *IEEE Transactions on Automatic Control*, 58(6):1613–1619.

[19] Yor, M. 1992. On some exponential functionals of Brownian motion. *Advances In Applied Probability*, 24:509–531.

VITA

Josh McDonald was born in Newbern, Tennessee on April 28, 1985, the son of Tommy and LonNita McDonald. After finishing high school in 2003, he enrolled in classes at Dyersburg State Community College before moving on to Austin Peay State University in Clarksville, Tennessee. There, he earned his Bachelor of Science degree with a major in mathematics and statistics in December 2007. He then enrolled in the graduate program of the H. Milton Stewart School of Industrial & Systems Engineering at the Georgia Institute of Technology in August 2009, and he defended his Ph.D. dissertation in August 2015.